Unit Trust Performance Metrics: A Comparison of Value, Growth, and Ethical Fund.

Abdel Said

A thesis submitted to the Faculty of Business and Law of the

University of the West of England for the degree of

DOCTOR OF PHILOSOPHY

March 2022

Acknowledgements

I would like to acknowledge the support given to me by numerous people throughout the entire doctorate process. My sincere gratitude goes out to my supervisor, Professor Cherif Guermat, who has provided me with a great deal of support and endless inspiration. I am very fortunate and appreciative to have had his ongoing support and guidance with this work. I am also grateful to Dr. Vasco Vendrame for his continuous support. I wish to thank the Graduate School at UWE for funding my PhD.

This thesis is dedicated to my father Rashed Said. Although unfortunately he left me too early, his encouragement and kindness will stay with me always. I thank my dear mother, Emilia Amya, for her unconditional love, support, patience and consistent motivation.

I thank my brothers, Ismail, and Bare, for their support, and my sisters Hamsa and Ansam for believing in me and for giving me all of the love and support I needed throughout this project.

A special thanks to my niece Yara, for giving me the hope and courage to push myself further than I ever thought I could.

Finally I thank my dear friends Jay Bond, Viki Kostadinova, and Pablo Gomez for being so wonderful and for making me smile during even my darkest days.

Abstract

Empirical evidence from the mutual fund industry suggests that on average, mutual fund managers do not possess stock selection abilities. Yet, there is always the possibility that traditional performance measures are not sufficient to detect such skill. A major challenge in performance measurement is to identify an appropriate benchmark against which fund performance is evaluated. Once this has been achieved, a further challenge regards whether investors can successfully choose funds based on an ex-ante strategy to generate positive abnormal returns in the future. The main objective of this thesis is to comprehensively explore the performance of UK-equity unit trusts, with a particular focus on ethical unit trusts. After noting the many caveats in fund performance measurement, we examine whether unit trust managers are able to select stocks and generate positive style-adjusted performance. Specifically, we employ a returns-based style analysis to identify funds' exposures to four key benchmark factors; small-value, small-growth, big-value, and big-growth. We then classify funds on the basis of their return attributes and draw a conclusion on the relationship between investment style and fund managers' stock selection skill. Next, we move to examine whether investors are able exploit an ex-ante investment style strategy to generate positive abnormal returns in the future. Using event study methodologies, we investigate the profitability of investment strategies that could be achieved by systematically buying units in funds with specific investment style objectives over an investment horizon of a one to five-year period. Finally, we deal with the data mining bias to assess whether fund managers performance can be attributed to luck or skill.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

RBSA: Return Based Style Analysis

BHAR: Buy and Hold Abnormal Return

CAR: Cumulative Abnormal Return

CTAR: Calendar Time Abnormal Return

AUTIF: The Association of Unit Trust Investment Funds

IMA: Investment Management Association

AUM: Assets Under Management

UT: Unit Trust

OIECs: Open-Ended Investment Company

ESG: Environmental, social and governances

EIRIS: Ethical Investment Research Services

SEDOL: Stock Exchange Daily Official List

KIID: Key Investor Information Document

Chapter 1

Introduction

1.0. Introduction

In principle, the investment management industry plays an important role in stabilising the financial market, both through the provision of liquidity and via the function they play in the price mechanism of financial assets. Investment managers therefore contribute to market efficiency by pricing information correctly and channelling the savings of individuals towards the most efficient investments. The importance of the investment industry is not limited to capital market efficiency, it also has an important responsibility with regards to stewardship, engaging with the companies they invest in to maintain and enhance value for their clients.

In the past few decades mutual funds have become a popular investment vehicle among investors. This popularity is attributed to their ability to offer effective risk diversification, and their having access to a wide range of asset classes and investment strategies. These significant benefits are considered the cornerstone of mutual fund investing. However, investors are faced with the dilemma of whether to allocate their savings into active or passive funds. Typically, passive funds imply an investment strategy of buying-and-holding a given market index for a long-term investment horizon. Thus, market prices are always fairly set and there is no view on how the price may differ from the current market price in the future. In contrast, active funds tend to buy and sell their underlying investments more frequently. The rationale of active funds is that managers can add value for their investors and generate abnormal returns above a set of benchmarks by using private information and their stock picking skills. However, charges for managed funds tend to be a lot higher than for passive funds.

Theoretically, the Efficient Markets Hypothesis (EMH) states that market prices reflect all available information at any given time (Fama, 1965). The price of an asset includes all available information and active fund managers are on average unable to generate abnormal returns. However, Grossman and Stiglitz (1980) argue that even when the market is informationally efficient, expected abnormal return should not be zero, otherwise there would be no reward to collect and process costly information. Managers who have informational advantages should therefore receive compensation to promote information gathering activity.

Empirically, the vast majority of research in fund performance evaluation suggests that actively managed funds are, on average, unable to outperform passively managed funds, once transactions cost and management fees have been taken into account (i.e., Fletcher, 1997; Blake and Timmermann, 1998; Quigley and Sinquefield, 2000). Yet, the cross-sectional distribution of individual funds' performance indicates that some managers do have systematically higher risk-adjusted excess returns (Kosowski et al., 2006; Cuthbertson et al., 2008; Blake et al., 2014). These findings do not necessarily violate the efficient market hypothesis since active management is a zero-sum game, whereby, for example, the market return must equal the aggregate returns on the passive and active segments of the market. Given those passive investors earn precisely the market return, then active investors must, in aggregate, also earn the same average return (before fees) as the market (Sharpe, 1991). Thus, if some active managers earn positive risk-adjusted returns, others lose with equivalent magnitude. This would raise the question of whether it is possible to identify fund managers who can generate a significant abnormal performance, and if so, how investors can exploit this opportunity.

1.2. Research Background

Many performance evaluation techniques seek to identify the existence of skill amongst active fund managers. The standard approach relies on a multivariate regression to explain the variation of funds' returns. Thus, risk-factors are used to separate funds' performance into returns from common risk-factors or from the fund managers' skill (i.e., Jensen's alpha). Single or multi-factor models are commonly used as the standard benchmarks to evaluate funds' performance.

The issues with this approach include the lack of appropriate benchmarks to explain the cross-section of returns, and the models' inability to capture the time varying in either the factor loadings or funds weights (bad model problem). In this strand of literature, factor models that assume linear relationship between funds' return and risk-factors suffer methodological biases and significantly affect performance evaluation results. In practice, fund managers follow a wide variety of strategies and portfolio weights, and systematic risk characteristics vary across time because of active stock selection. Thus, it is unlikely for a single static model to capture the time varying risk factors across all funds (Brown and Goetzmann, 1997). Although cross-sectional rolling regression, conditional models, switching models or the adoption of Kalman filtering are potential remedies for such a problem, risk factors must match funds' investment holdings.

Besides the bad model problem of market equilibrium, the statistical inferences are only correct when the abnormal returns from the performance model are normally distributed. Typically, the significance of abnormal performance is tested using the standard parametric t-test, whose validity relies on the assumption of normal distribution. However, a fund's return often departs from normality (i.e., Cuthbertson et al., 2012; Ornelas et al., 2012), simply because of the fact that its holdings contain assets with similar characteristics. Moreover, even if individual fund returns are normally distributed, the cross-sectional distribution of the abnormal returns may be non-normal due to industry clustering. The violation of this assumption leads to inefficient estimates and a low statistical power to detect abnormal performance. Furthermore, tests of funds' abnormal performance often ignore the fact that some funds over-perform (under-perform) because of good (bad) luck rather than good (poor) skill (see for example, Kosowski et al., 2006; Cuthbertson et al., 2008; Fama and French, 2010; Busse et al., 2010; Blake et al., 2014). This motivates the use of a bootstrap approach to provide statistical validity to the performance metrics and ascertain whether the source of abnormal performance is due to skill or luck.

The multivariate approach provides a partial answer to the question of whether fund managers add value for their investors. A central issue within this debate is whether investors can successfully choose a fund based on an ex-ante strategy that will generate positive abnormal returns in the future. One way to tackle this question is to implement the recursive portfolio approach. According to this method, funds are sorted into equal or value weighted portfolios based on decile ranking of performance attributes with periodic rebalancing. Thereafter, post ranking returns are used to evaluate future performance.

However, performance attributes may not always be explicitly specified. There are many factors that might influence a funds' sources of return over time (i.e., investment style, size, age, past performance, charges, net inflow, managers and fund family characteristics). As the number of possible funds sorting rules to form an ex-ante strategy increase, the post-sorting portfolio returns are most likely to suffer from non-normal idiosyncratic risk, whereby the predicted performance may differ significantly from past returns as a result of random attributes. Hence a bootstrap approach may be required to capture the cross-sectional correlation in post-sorting portfolio returns for valid inferences (Kosowski et al., 2006; Cuthbertson et al., 2008; Javier, 2013; Blake et al., 2014).

Another important issue is the time horizon over which performance measures are calculated. The recursive portfolio formation approach requires periodic rebalancing toward the funds' attribute target. This generates the holding period returns of the ex-ante strategy. Therefore, the predicted performance is normally computed over various holding periods ranging from one month to five years. For example, Quigley and Sinquefield (2000), Blake and Timmermann (1998), Tonks (2005), Cuthbertson et al. (2008) and Fletcher (2015) investigate the persistence in past performance over a multi-period time horizon after portfolio formation. However, periodic rebalancing inherent in recursive portfolio formation might introduce false inferences. In particular, the predicted abnormal return may not correspond to the true returns that investors accumulate by the end of the holding period. Furthermore, investor's end wealth must take into account any rebalancing costs between funds such as search costs, load fees, administrative and advisory fees (Cuthbertson et al., 2016).

1.3. Research Questions

Q1: What investment style is adopted across UK-equity funds? Is style investing profitable? Do ethical funds invest differently than their conventional counterparts and do they pay a higher price for their ethical consideration?

To address these questions, we examine fund managers' stock selection in the context of return-based style analysis. By decomposing funds' return into size and growth-value dimensions, we explore whether funds' performance differs across styles and scrutinise fund managers' ability to generate abnormal return on a style-adjusted basis. In the RBSA original work proposed by Sharpe (1992), a fund's return is regressed on 12 indices of major asset classes that covers the investment universe available to fund managers. The regression coefficients are restricted to non-negative value and add up to one. Thus, the estimated coefficients provide meaningful descriptors of a fund's investment style (weight), and the residuals determining fund managers' stock selection skill. In practice, managers of domestic equity funds often break the domestic equity universe down into size and value growth dimensions. Therefore, these style dimensions are commonly used to describe products offered by mutual funds and to assess fund managers' performance and stock selection behaviour (Chan et al., 2002). Accordingly, the Fama and French three-factor approaches are widely accepted both theoretically and empirically as representative factors of funds' investment style (Chan et al., 2002; Brookfield et al. 2014).

In our implementation of the RBSA approach, we chose four style mimicking portfolios to reflect the returns on size and value-growth factors. By separating the Fama and French three-factor approaches, for example small-minus-big (SMB) and high-minus-low (HML) into small-growth, small-value, big-growth and big-value, we account for the growth/value effect across small and big stocks, whilst also retaining the link to the empirical asset pricing literature in identifying risk factors. Given the constraints on the regression coefficients, the factor loadings are estimated by nonlinear optimization procedures such as the BFGS (Broyden, Fletcher, Goldfarb, and Shannon) method. Furthermore, Fama and Macbeth (1973) rolling regression is employed over a 36-month window to capture time-variation in factor loadings. Funds' returns are thereafter sorted into four stylized portfolios based on their factor exposure using either a yearly rebalancing strategy or the average value of factor exposure over the whole sample period. Each stylized portfolio is assigned a style benchmark (mean-variance efficient portfolio) based on a style estimation before any style-based performance evaluation is carried out, using both a style benchmark and a general market index (FTSE100). A similar procedure is applied to determine whether there is any difference in ethical funds' performance across investment styles. We also compare the performance of stylized ethical portfolios against their conventional peers using investment style as the matching criteria.

Q2: If fund managers produce alpha attributed to fund's investment style, can investors exploit an ex-ante investment style strategy and how should they frame their investment horizon between competing strategies?

While measuring the impact of fund investment style on performance might seem straightforward, a crucial issue is whether investors are able exploit a successful ex-ante investment style strategy. The common practice utilised in measuring investors' style/risk-adjusted performance is via the creation of a recursive portfolio which allows for a direct assessment of investors' terminal wealth as well as the statistical significance of such an investment strategy. Usually this is carried out in the context of post-sort performance alphas and using either the event-time or the calendar time approach. However, there is no general agreement on which approach is more appropriate; each has its pros and cons. Schultz (2003) argues that when the events under consideration are correlated (i.e., when investment style rotation has occurred because of market conditions, or window dressing has occurred before the fiscal year-end), the event time approach introduces false inferences and often leads to a rejection of the null hypnosis of no abnormal returns. Furthermore, Fama (1998) and Mitchell and Stafford (2000) suggest that the calendar time approach is more powerful in capturing the expected returns, thus eliminating the bad model problem. In contrast to the event time approach, the calendar time approach involves rebalancing the portfolio at the start of every month. This approach is inaccurate in capturing the true return from a buy-and-hold strategy over the investment holding period (Loughran and Ritter, 2000). Liu and Strong (2008) show that evaluating the performance of investment style by formulating a single-period portfolio return over a multi-period holding horizon is misleading and produces biased statistical inferences. Another problem is that the regression factor loadings are assumed to be stationary over the period of the study, even though funds are, every month, added or excluded from the calendar time portfolio. Thus, the regression suffers from heteroscedasticity and standard errors are biased and inconsistent, which in turn leads to invalid hypothesis testing and confidence intervals.

Since the core objective is to examine the feasibility of investors profiting from the pursuit of a successful extant investment style strategy, our preferred method to measuring the abnormal performance of the aggregate returns of funds' investment style is the Event Time Buy and Hold Abnormal Return (BHAR) approach. Here the abnormal return is calculated according to the principles set out in Liu and Strong (2008). Thus, we measure the abnormal performance that could be achieved by systematically buying units in funds with a specific investment style over a one to five year investment horizon. This approach has the advantage of measuring investors' true returns on the underlying investment strategy. It maintains the buy-and-hold property, and therefore properly indicates a typical investor's terminal wealth from adopting such a strategy. We also report the results from the Cumulative Abnormal Return (CAR) method for completeness. The sample period under study saw severe economic fluctuation, including both the global financial crisis of 2007 and the European sovereign debt crisis of 2011. This economic fluctuation would strongly challenge the normality assumption of mutual funds' returns, and hence induce miss-specified inferences of the standard test statistics. Accordingly, we apply a sophisticated wild-adjusted bootstrap approach to allow for appropriate statistical inferences in the presence of non-normal fund returns. Finally, in the spirit of the recommendations of Fama (1998), and Brav et al. (2000), the calendar time abnormal return is also considered to account for cross sectional correlation of abnormal returns. The abnormal return is calculated as the mean abnormal time-series of event funds' portfolio returns over a five-year holding period. The statistical inferences are robust to heteroscedasticity bias; more specifically the standard errors are robust as a result of using the OLS with White's correction and Gregory et al.'s (2010) Feasible GLS technique.

Q3: Using style/factors-adjusted benchmark, how much of any fund manager's performance is due to luck (good or bad) and how much is due to skill (good or bad) and does an investing style matter to fund performance?

The standard approach for evaluating fund manager skill is to measure the significance of abnormal returns produced by an appropriate benchmark model. However, there are two main issues with such an approach. First, fund's returns exhibit non-normal distribution. Second, the test ignores the fact that abnormal performance might be largely due to luck. Therefore, tests of fund's abnormal performance may give misleading inferences and little economic value to investors. Instead, Kosowski et al. (2006), and Fama and French (2010) propose bootstrap approache to examine whether abnormal performance is due to managerial good/bad skill or genuine good/bad luck. The basic concept of these approaches is to compare the performance distribution of the mutual fund against a simulated luck distribution with zero abnormal performance but that has the same statistical properties as the actual fund. If the actual fund performance distribution exceeds the simulated distribution, then it is regarded as evidence of genuine good/bad skill.

Following this line of reasoning, it is important to determine whether fund managers abnormal performance is attributed to luck or skill. This thesis aims to extend the luck versus skills debate using factor models and style adjusted benchmarks. To achieve this, we implement the baseline (Kosowski et al., 2006) bootstrap technique and we also introduce the skewness-adjusted and kurtosis preserving wild bootstrap. The main advantage of the wild-adjusted bootstrap is that the luck distribution is constructed in such a way that mimics the four moments of the true fund's returns distribution. Since investors are interested in the distribution of their terminal wealth from active fund investment; for example in terms of mean abnormal performance, skewness, and kurtosis, we advocate the use of the wild-adjusted bootstrap as it has better statistical properties in distinguishing skill from luck in fund performance. We conduct the analysis for both gross and net returns. Thus, we examine whether funds managers can pick stocks well enough to cover their operating costs and management fees. We also investigate whether manager skills differ across different investment styles. By doing so, we account for any homogeneous risk across funds, which might not otherwise be captured by the benchmark models.

1.4. Contribution to Knowledge

To measure abnormal return, it is necessary to identify an appropriate benchmark for expected return. For example, Angelidis et al. (2013), and Mateus et al. (2016) argue that commonly used factor models are unable to identify fund's managerial skill if the fund's style characteristics differ from those of the benchmark index. Thus, mean-variance efficient portfolios might be achieved by constructing a style-adjusted benchmark. Such an approach will therefore be effective in evaluating fund's abnormal return. There are two important implications to our style analysis in the UK mutual funds literature. First, we use a survivorship bias-free dataset of UK-equity funds to assess fund managers' style-adjusted performance. We augment the commonly used factor models by creating a style-adjusted benchmark that quantifies the performance more effectively than the general market index. Hence, we enable improved inference in the evaluation of mutual fund performance. Second, we provide evidence that the size and value-growth investment style dimensions perform the best in fund's performance evaluation measurements. We do so by comparing the extent to which fund's returns track its style-adjust benchmark more closely than the general market index. Thus, we advocate the use of a style-adjusted benchmark as a standard practice in mutual fund performance evaluation.

Our discussion surrounding investors' ability to exploit a successful ex-ante investment style strategy is quite novel to fund performance literature and has an appealing feature for fund investors. In particular, we discuss practical issues in implementing an ex-ante investment style strategy and in measuring investor's terminal wealth through the adoption of such a strategy. This is achieved by focusing on the recursive portfolio formation and the methods of measuring abnormal returns. Hence this work can serve as a significant guideline for fund investors' attitudes towards investment style, investment horizons and the frequency of portfolio rebalancing. For example, if abnormal performance is found to be significant at a specific time horizon, then it may represent an exploitable strategy for investors. By comparing the short and long-run abnormal performance defined by both the event and calendar time, we contribute to the discussion of whether performance is sensitive to the choice of empirical method and the investment horizon. We also provide empirical evidence on ethical fund investors' experience and whether they pay a price for their ethical considerations compared to their conventional counterparts.

This thesis also contributes to the fund performance luck versus skill debate in two ways. First, by comparing the baseline and the wild-adjusted bootstrap from benchmarks models, any selectivity skills that fund managers might possess are robust with respect to the variability of the bootstrap methods, and benchmark models. Therefore, we provide improved inferences in the evaluation of mutual fund performance, accounting for the non-normality and heteroscedasticity of individual mutual fund returns. Second, our result has an important implication for investors and the mutual fund industry. It demonstrates that the common practice of ranking funds according to their past performance (i.e., Morningstar's five–star rating system for mutual funds) gives little information about a fund manager's stock picking talent or a fund's future performance.

1.5. Structure of the Thesis

This thesis is structured as follows. Chapter 2 provides the core theoretical and empirical literature review of the mutual funds' performance measurement. In particular, it discusses various performance evaluation models, and different fund's investment styles. It also reviews the literature findings in relation to abnormal return measurements in short-run and long-run event studies. Finally, it presents the literature on persistence and skills versus luck in mutual fund performance measurement.

Chapter 3 provides a detailed description of the data employed in this thesis as well as data definitions and sources. It also provides descriptive statistics of the unit trust returns, and benchmark factor portfolios.

Chapter 4 presents the methodologies applied in this research to evaluate mutual fund performance. The chapter starts with a description of the return-based style analysis approach. This is followed by the construction of the stylized portfolios based on continuous changing style and dominant (constant) style. Thereafter, event time portfolio formation is introduced along with the BHAR and CAR measures. Next, we present the calendar time methodology and the calendar time portfolio construction. Finally, we present the simulation-based procedures including the baseline and the wild-adjusted bootstrap.

Chapter 5 presents the results of style/risk adjusted performance. Then, a comparison between the various specifications of the model is undertaken to determine which benchmark provides a better explanation of the cross-sectional variation in funds returns. It also presents the empirical results of ethical funds' style and their performance. Finally, we demonstrate how the returns-based style analysis (RBSA) is applied in practice by analysing two randomly chosen individual UK equity funds.

Chapter 6 discusses the results of conventional and ethical funds' investment styles using the BHAR and CAR methods. In particular, we discuss the propensity of these strategies to generate statistically and economically significant abnormal returns for an investment horizon of one to five years. It also shows the statistical significance procedures of both the skewness adjusted and the kurtosis preserved wild bootstraps.

Chapter 7 presents the results of the stylized calendar time portfolio over a five-year holding period for both conventional and ethical funds. The statistical inferences are robust to heteroscedasticity bias using the OLS with White's (1980) correction and Gregory et al.'s (2010) Feasible GLS technique.

Chapter 8 discusses funds' performance after controlling for the luck factor using the baseline and wild-adjusted bootstrap. It also discusses whether skill is concentrated within a certain investment style.

Chapter 9 concludes the thesis and re-addresses the research questions to determine whether funds managers add value to their investors. The chapter also highlights the main limitations of the research and provides recommendations for future research.

Chapter 2

Literature Review

2.0. Introduction

The performance evaluation of a managed fund has received a great deal of attention in the academic literature and among practitioners. A key question of interest is whether active fund managers can outperform passive managers and add value for investors. To address this question, various measures have been proposed to evaluate active funds' performance. This chapter presents a review of the existing literature on unit trust performance and fund managers stock-selection skill. We handpicked studies that we thought provide an insight into active funds' performance and its content is directly relevant to this thesis. The review starts by shedding light on performance measures in the context of regression models. This section includes a discussion of single and multi-factor models for measuring the stock-picking abilities of active fund managers. We then turn to long-horizon abnormal performance. Here we discuss event studies and associated tests of the significance of abnormal performance. Finally, the literature findings in relation to fund performance and investment style, skill versus luck in performance and ethical fund performance are reported.

2.1. Performance Measures:

2.1.1. Single-Factor Model

The earlier studies on mutual fund performance used the capital asset pricing model (CAPM) to generate expected returns. To examine whether fund managers add value, Jensen (1968) introduced the standard technique of alpha performance measurement. The intercept (alpha) is measured by regressing fund excess returns against excess return on the market portfolio. Thus, the Jensen's alpha assesses the fund's level of abnormal performance compared to what the CAPM would predict, for example a positive alpha is considered as evidence of fund managers' stock selection skill. However, the validity of Jensen's alpha depends on the legitimacy of the CAPM and its accuracy in explaining expected returns. Indeed, Roll (1978) shows that it is impossible to observe the market portfolio, which in theory is an efficient and fully diversified portfolio that contains all the risky assets in the economy. This casts doubt over whether the market index is a good proxy of the unobserved mean-variance efficient portfolio. Furthermore, performance measures based on the CAPM have created a heated debate in finance literature, specifically with the discovery of the market anomalies in the early 1980s. For example, Banz (1981) and Reinganum (1981) show that small stocks earn higher returns compared to large stocks, and Basu (1983) demonstrates that the book-to-market ratio have explanatory power for the cross-sectional variation in stock returns. These anomalies imply that a single risk factor model does not provide a complete description of the cross-section of fund returns.

2.1.2. Multi-Factor Models

Confronted with the empirical failures of the CAPM, Fama and French (1993) propose two additional factors formed on market capitalization and book to market ratio to accommodate the anomalies that the CAPM failed to capture. The size factor captures the excess returns on a portfolio of small stocks relative to a portfolio of large stocks. Similarly, the book-to-market factor captures the excess returns on a portfolio of value stocks relative to a portfolio of growth stocks. These two factors can be viewed as mimicking portfolios and of acting as a proxy for sensitivity to common risk factors. Although they have been successful and widely used in empirical research, the underlying risks are still unclear (Fama and French, 1996). For example, what economic risk factors are small and value stocks compensated for? Banz (1981) suggests that small stocks are normally illiquid and subject to higher volatility than big stocks. Lakonishock et al. (1994) show that value premium is attributed to irrational pricing of those stocks (i.e., investors overconfidence). On the other hand, Fama and French (1996) suggest that value premium is related to distress risk. For example, stocks with poor expected performance usually have low share prices hence high book-to-market ratio. However, the Fama and French three-factor model is not free from anomalies. Jegadeesh and Titman (1993) show that stocks with high returns in the past 12 months tend to have high future returns. This suggests that stocks with positive previous returns will continue to generate positive returns and should be considered as an investment prospect. Carhart (1997) extends the Fama and French three-factor by adding a momentum factor as an extra variable to capture the cross-section of expected return. Therefore, fund managers should not be rewarded for following a strategy such as this which can be easily implemented based on publicly observable information.In a more recent study, Fama and French (2015) propose a five-factor model by augmenting the three-factor model with two extra factors that capture the return premiums associated with profitability and investment. The dividend discount model is used to substantiate the addition of the two mimicking factors on asset returns. The profitability and investment factors are formed based on operating profitability robust minus weak (RMW) or investment conservative minus aggressive (CMA). The RMW and CMA can be interpreted as averages of profitability and investment factors for small and big stocks. Fama and French (2015) argue that their five-factor model is superior to the three-factor model in explaining variations in average monthly stock returns of NYSE over the period 1963-2012.

2.1.3. Style Analysis

Sharpe (1992) notes that fund managers are restricted to invest in predefined asset classes and the only discretion allowed is to select shares within each asset class. Accordingly, he proposes a 12-asset-class model in order to fully capture the variation of fund returns. In this approach, fund performance is compared to the performance of commercial indexes that are mutually exclusive, something that can be easily implemented by investors. The Sharpe (1992) regression coefficients are also constrained to non-negative and add up to unity. Therefore, the coefficients not only signify the exposure to different asset-classes but can also be interpreted as fund weights. This helps to determine the style of any fund from its betas exposures and forms the basis of Return Based Style Analysis (RBSA). Daniel et al. (1997) and Wermers (2004) decompose fund returns based on the characteristics of the fund's stock holdings instead of identifying fund's style exposure from the sensitivities of its return to factors. This is achieved by observing the stock holdings of the fund, then designing benchmarks based on characteristics matching criteria. Fund's returns can therefore be evaluated in terms of whether superior performance is generated from manager's stock-picking skills. This approach has become known as characteristics-based style analysis. By comparing the return-based and the characteristics-based style analysis, Chan et al. (2002) show that the two approaches provide similar interpretations of a fund's investment style. The returns-based method is as powerful as the characteristics-based approach in explaining the cross-sectional behaviour of fund's returns.

The literature concerning factors that capture the variation of fund's returns is both extensive and compelling. For example, some researchers view the three-factor model as a variant of Sharpe's asset classes model. They argue that the size and value dimensions are reflective of two main style classifications that are widely implemented by domestic equity funds (Bassett& Chen 2001). Furthermore, index providers such as Standard & Poor's, Russell, MSCI, and Morningstar benchmark fund's return in a way that reflects the size and value dimensions approach. Fund managers' stock selection skill may therefore be detected more clearly when fund returns are matched with benchmarks that mimic these underling strategies (Chan, et al., 2002; Ben Dor et al., 2003). The three-factor model therefore remains the prevalent factor benchmarking model present in the academic literature.

However, Huij and Verbeek (2009) show that the way a risk factor is constructed can significantly affect performance measurement, specifically in relation to how risk factors handle fees, commissions, taxes, trading restrictions, and dividends. Huij and Verbeek argue that, in the real world, returns on passively managed indices are likely to be lower than those forecasted by theoretical risk factors. Similarly, Blitz and Huij (2012) conclude that a hypothetical passive investment index is not suitable for evaluating the performance of actively managed funds. Yet, the most damaging claim against the Fama and French three-factor model is the presence of abnormal performance in passive portfolios used as benchmarks in the model. For example, Cremers et al. (2012) show that regressing the S&P 500 (Russell 2000) Growth index on factor models results in a statistically significant positive (negative) alpha. They argue that the failure of the three-factor model is due to the methodology of constructing the risk factors. For example, the size factor assigns equal weights for both value and growth stocks, resulting in an overweighting of small value stocks. Since small value stocks have outperformed other stocks in the past, such a tilt in weight exaggerates the returns on SMB factor causing misleading inferences. Similarly, Angelidis et al. (2013) argue that factor proxies themselves have significant alphas which often bias the performance evaluation procedure. This is of crucial importance to investors who want to invest in active funds. For example, if an active fund manager is holding a portfolio that mimics the S&P 500, then investor interpretations of this fund manager's skill is biased.

In response to these issues, two strands of recent academic literature have emerged. The first strand of the literature concentrates on supplementing the model with factors or indexes to provide better explanation of the cross-sectional variation in fund returns and an appropriate indicator of fund managers' skill. Among these studies, Pastor and Stambaugh (2001) demonstrate that including information in returns on non-benchmark passive assets helps to reduce sampling errors in the regression. For example, the expected return on a fund with technology investment exposure is better described by the three-factor model augmented with a passive technology index. Moreno and Rodriguez (2009) suggest that a coskewness factor in mutual fund performance evaluation is economically and statistically significant. Huang et al. (2012) show that liquidity is a proxy of systematic risk. Henter et al. (2014) argue that using gross average return on all mutual funds (active peer benchmark) in addition to market, size, and value risk factors can significantly improve the power of the factor model.

The second strand of literature tries to identify the sources of future fund performance using fund specific characteristics, such as size, past performance, stock holding, management fees, manager characteristics, and internal governance. In this approach, researchers examine whether abnormal performance can be identified ex-ante according to the funds' characteristics or attributes under consideration. Typically, a multivariate regression is used to capture funds' properties (attributes), where Fama and MacBeth (1973) cross-section rolling regression is commonly employed to obtain estimates of time varying betas. Funds are then ranked periodically into decile portfolios based on their returns over the preceding holding period. Finally, the post ranking performance over the subsequent period is benchmarked using a factor model to determine the magnitude of abnormal performance.

However, this approach has its own shortcomings, as it suffers from difficulties in pinpointing contributing factors and there is no guarantee of its continuation. For example, numerous studies conclude that the performance of actively managed funds can be attributed to past performance (persistence in performance). But then Berk and Green (2004) show that fund performance is a decreasing return to scale. In other words, fund managers' ability to produce alpha declines as the size of a fund increases. Therefore, a fund's past performance could predict returns only if the fund's net inflows are considered. Clearly there are a large number of factors that might influence future performance, and the more contributing factors are considered the more likely their effect is cancelled out or reduced.

2.2. Event Studies

Kothari and Warner (2001) alert researchers to the low power of factor models to detect abnormal performance. The measurement of abnormal returns is grappled with via the choice of suitable benchmarks. Besides, the intercept (alpha) of factor models measures performance over several years, whereas investors are normally interested in performance measures that range from three to five years. Event studies offer an alternative approach to measure abnormal performance and could potentially alleviate these issues.

The literature of performance measurement following a corporate event is vast and considers many different issues such as the design and statistical properties of event study methods. There are two main methodologies commonly used in event studies, namely, the event-time approach and the calendar-time approach. In the event-time approach, researchers have to choose between the buy and hold abnormal return (BHAR), and the cumulative abnormal return (CAR). The main difference between the two methodologies is that the BHAR uses geometric returns while the CAR employs arithmetic return in calculating the overall return over the event period of interest. Therefore, the BHAR procedure allows the compounding of returns over a longer period, which better measure investors' experience as opposed to the periodic rebalancing entailed in the CAR approach (Barber and Lyon, 1997; and Lyon, et al., 1999).

Studies investigating the long-run performance are highly controversial and far from settled. Obviously, the joint hypothesis problem remains at the centre of this debate, whereby It is not possible to measure abnormal returns without expected returns predicted by pricing models. Most commonly, two notions are adopted for expected returns including factor models, and reference portfolio or a control firm approach as a benchmark for measuring abnormal returns. For example, the event firms' abnormal performance can be measured against the abnormal performance of the size, book to market value, and momentum matched reference portfolio of non-event firms. Kothari and Warner (1997) and Lyon et al. (1999) show that the event-time abnormal performance is likely to be subject to biases arising from new listings, and the rebalancing of benchmark portfolios. For example, the return of a reference portfolio is commonly calculated using periodic rebalancing, while the returns of the event firms are computed without rebalancing. Thus, researchers must be cautious when constructing the reference portfolio in order to mitigate these biases (Lyon et al., 1999). However, the matching procedure is valid if and only if the event firms and non-event firms have a similar expected return. This condition is highly volatile when event firms experience extreme pre-event performance, particularly if the event is anticipated or non-random (Kothari and Warner, 2006). A further problem is that the event-time portfolio does not represent an investable opportunity since the total number of event firms is not known in advance (Fama, 1998).

Many Researchers reject the event-time approaches in favour of the calendar time portfolio (Fama, 1998; Mitchell and Stafford, 2000; and Brav et al., 2000). The calendar time portfolio eliminates the issue of cross-sectional dependence of returns, assuming that the market is efficient. However, Loughran and Ritter (2000) criticise the calendar time approach since it ignores the timing of the managerial decisions, and falsely leads to a conclusion that is consistent with market efficiency. In practice, managers time their corporate events to exploit mispricing. For example, managers are more likely to raise additional capital through secondary offering following price rises. Therefore, the event time approach is preferable to the calendar time as it weights event firms equally rather than weighting each time-period equally. Furthermore, Lyon, Barber and Tsai (1999) point to the fact that the number of event firms vary in the calendar time, since event firms are periodically added or removed from the calendar time portfolio. As a result, the regression is obviously heteroskedastic, and coefficients might be biased. Although Fama (1998) notes the heteroskedasticity issue in a calendar time portfolio, he strongly advocates the use of calendar time over the event time approach. Their argument is that the bad-model problem is less significant in the calendar-time approach and that most anomalies disappear when performance is measured using calendar time.

2.2.1. Statistical Power of the Tests

In addition to the bad model problem, there are several issues concerning the aggregation of event firms and the test of the statistical significance of abnormal returns that are crucially important in the event studies. There is significant empirical evidence in the existing literature that the distribution of long-run returns deviates from the normality assumption that underlies many statistical tests. Barber and Lyon (1997) show that the return of an event firms' portfolio exhibits fat-tail distribution (kurtosis) because of large price jumps around the event period. Furthermore, the return tends to be skewed to the right, so the t-statistic is asymmetric. This phenomenon is known as the stylized fact, where the long-run performance distribution is restricted to -100% and unrestricted on the upside (Barber and Lyon, 1997). Although, the mean returns revert to zero as the sample size increases in accordance with the central limit theorem, the skewness bias arises because the lack of independence among event firms (Brav, 2000; and Mitchell and Stafford, 2000). For example, pre-event performance, industry clustering, and firms with overlapping returns seem to cause severe misspecification and hence result in treacherous inferences. Cross-sectional correlation in firms' residuals represent a major challenge in event-time studies, specifically in the long-run. Mitchell and Stafford (2000) note that the calendar-time approach can mitigate this puzzling issue. However, they both recognise that the cross-sectional correlation cannot be ignored over the long run.

The focus of research shifts towards using non-parametric approaches to deal with non-normally distributed returns in event studies. Lyon et al. (1999) propose a bootstrap approach with a skewness-adjusted t-statistic to account for the non-normality in long-term abnormal returns. This involves adjusting the t-statistic of the abnormal returns using Johnson’s (1978) approach, before constructing an approximate empirical distribution from the original sample of abnormal returns, typically between 1000 and 10,000 resamples from the parent distribution. The skewness-adjusted t-statistic of the parent distribution is then compared to the bootstrap’s empirical distribution to determine whether the abnormal performance is statistically different from zero. Ikenberry et al. (1995) perform a pseudo-portfolio approach in order to control for the skewness bias in tests of long-run abnormal returns. Each event firm is matched with a randomly chosen non-event firm with similar characteristics in terms of size and book-to-market value at the time of the event. The process is repeated until each event firm is represented by the control firm in the pseudo-portfolio, before the abnormal return is computed from the pseudo-portfolio. This technique is repeated one thousand times to generate the empirical distribution. Finally, a five percent cut-off point of the empirical distribution is used to conclude whether the abnormal return is statistically significant. Both Lyon et al. (1999) and Ikenberry et al. (1995) conclude that the BHAR test is well specified. The bootstrap procedure together with careful benchmark portfolio formation correct for the non-normality generated by listing, rebalancing and skewness bias. Gregory et al. (2010) control for the heteroscedasticity and kurtosis bias in tests of long-run abnormal returns by applying a skewness-adjusted t-statistic and kurtosis preserving wild bootstrap. Their approach is intended to ensure that the empirical distribution mimics the parent distribution as closely as possible.

However, the argument against the assumption of cross-sectional independence of returns remains. In fact, several studies find that the cross-sectional dependence problem is more severe with BHAR than the CAR approach, due to the compounding property inherent in the BHAR. Thus, a minor error or miss-estimations in the event-period will lead to false inferences of abnormal performance over the long run. Jegadeesh and Karceski (2009) argue that the bootstrap approach cannot addresses the independence assumption since event firms are chosen in a non-random manner. Subsequently, they advocate the application of a correlation and heteroskedasticity-consistent test. They use both Hansen and Hodrick (1980) and generalize White's heteroskedasticity-consistent estimator to obtain unbiased t-statistics. They conclude that their tests provide an intuitive way to correct for heteroskedasticity and autocorrelation, specifically when the sample returns for event firms are clustered around a specific industry or contain overlapping returns.

Mitchell and Stafford (2000) strongly advocate the application of the calendar time approach to the measurement of long-run performance. They acknowledge that the regression may suffer from heteroscedasticity problems, and that therefore the ordinary least squares estimator is inefficient. A possible heteroscedasticity problem may arise from the fact that the number of event firms in the calendar time portfolio varies over time, and residual variance is likely to vary as well. In addition to restricting the minimum number of event firms in the calendar time portfolio at each point in time, they also apply the weighted least squares method aiming to improve the statistical properties of abnormal returns. Similarly, Gregory et al. (2010) recognise the heteroscedasticity problem in the calendar time portfolio. They propose a feasible generalized least squares method, instead of the simple weighted least squares method.

Their argument is that the effect on the residual variance from changing the number of firms in the event portfolio is unknown. Therefore, they assume that the heteroscedasticity takes the form of a linear function of the number of firms at each point in time. The residual variance is then neutralized by transforming the regression using an estimate of the variance. They conclude that the Feasible GLS delivers similar standard errors as in the OLS with robust White's variance estimators but has a better adjusted-R-square.

2.2.2. Simulation Studies on Long-horizon Event Studies

Barber and Lyon (1997) investigate the power and statistical properties of event studies to detect long-run abnormal returns. The study examines one-year CAR and BHAR returns by randomly selecting 10,000 monthly observations of 100 firms listed on the New York Stock Exchange. For each firm, the 12-month CAR and BHAR is calculated using the CRSP NYSE/AMEX/NASDAQ equally weighted benchmark index. Firms are then ranked into 100 portfolios according to their abnormal performance. When comparing the mean difference between CAR and BHAR, if the portfolio returns are more volatile than the returns on the benchmark, the magnitude of CAR is significantly higher than the BHAR. However, the difference is zero when the 12-month BHAR approaches the 28% threshold and become negative once the 12-month BHAR exceeds the 28% threshold. Furthermore, when comparing the mean and median of the BHAR for 200,000 random observations, the median and the mean of 12-month BHAR are -0.48% and -7.23%, respectively. This implies that the BHAR distribution is positively skewed and results in negative bias in the standard t-statistic. In contrast, the skewness is less pronounced under the CAR approach. In addition, Barber and Lyon also find that both BHAR and CAR are subject to rebalancing bias, and new listing bias. The rebalancing bias arises from a periodic rebalancing of securities that constitute the benchmark index. Consequently, the benchmark index is overstated, and this causes a negative bias in the long-run abnormal returns. New listing bias is also triggered as a result of newly listed firms after the event month. Therefore, if newly listed firms underperform the market as documented in Ritter (1991), then the long-run abnormal return is positively biased. However, Barber and Lyon (1997) strongly recommend using the BHAR over the CAR when evaluating long-run abnormal performance, since the CAR ignores the compounding effect in calculating long-term returns.

Lyon et al. (1999) carry out an elaborate analysis of long-run performance for firms listed on the NYSE, AMEX, and NASDAQ over the period from 1973 to 1994. The benchmark for expected returns is formed using either a reference portfolio approach or a control firm approach, based on the size and book-to-market of all firms. The one- three- and-five-year returns for the reference portfolio is calculated in two ways. First, the mean monthly returns for each of the size and book-to-market portfolios is calculated, then the returns are compounded over the investment horizon. This formation method is commonly used among researchers and requires periodic rebalancing to maintain an equally weighted portfolio. In the second method, Lyon et al compound the returns of firms that constitute the size and book-to-market portfolios, before calculating the mean returns across firms. This guarantees that the sample mean will be approximately normal (i.e., the population mean is guaranteed to be zero by construction). The main advantage of this procedure is that it eliminates the new listing bias and rebalancing bias. However, the benchmark portfolio might be partially subject to rebalancing bias, since the proceeds of the delisting firms are invested equally across firms constituting the reference portfolio. The authors also constructed the benchmark for expected returns using a control firm approach. Each firm is matched with a randomly chosen firm whose characteristics in terms of size or book-to-market ratio are closest to that of the sample firm. The BHAR is calculated over a one, three, and five-year holding period of 1,000 random samples of 200 event months without replacement. The test specification reveals the following: under both reference portfolios, the standard t-statistic is negatively biased. However, the scale of negative bias is less pronounced under the buy and hold reference portfolio (second formation method), and this can be attributed to the positive skewness of the long-run abnormal performance. When the t-statistic is adjusted using Johnson's (1978) skewness correction approach, the result demonstrates marginally less negative bias. However, the control firm approach, and the t-statistic skewness-adjusted bootstrap appear to be promising, and the misspecification seems to be mitigated. Similarly, when the critical value is computed using the pseudo-portfolio, the test is well-specified. Furthermore, Lyon et al examine the power of the test by introducing a constant level of abnormal return to the sampled firms. They conclude that the skewness-adjusted bootstrap and the empirical p-value derived from the abnormal performance of the pseudo-portfolio have a higher power of the resulting test statistic compared to the standard t-statistic.

All of the above statistical tests are conducted using random samples. Subsequently, they carry out similar tests using non-random samples based on size, book-to-market ratio, and pre-event performance. The results show severe negative bias across all test statistics, except for those generated using a control portfolio approach. In addition, they investigate the impact of the independence assumption caused by calendar clustering and overlapping returns. Their conclusion is that violating the independence assumption leads to estimation bias and mis-specified test statistics. Overall, Lyon et al. (1999) advocate for the use of BHAR to measure the true returns to investors. However, extreme caution is needed when constructing the reference portfolio, and the normality assumption can be corrected using either the skewness adjusted bootstrap or the empirical p-value of pseudo-portfolios.

Mitchell and Stafford (2000) provide further evidence on the statistical significance and power of the BHAR and CTAR to detect long-run abnormal returns. This study analysed the extent to which major marginal decisions (i.e., mergers, SEOs, and share repurchases) influence the long-run abnormal performance between 1958 and 1993 in the US stock market. Following Lyon et al.'s (1999) BHAR approach, the 3-year BHAR is calculated for each event firm using both value and equally weighted non-rebalanced 25 reference portfolios formed on size and book-to-market characteristics. To assess the statistical reliability of the skewness-adjusted bootstrap of the BHAR results, they simulate an empirical distribution under two assumptions. First, the reference portfolios perfectly capture the variation in sample firms. Second the empirical distribution mimics the true sample firms' distribution. Of course, the first assumption is problematic for all expected returns model, but it can be alleviated with a variety of model of expected returns.

The authors turn their attention to the second assumption, specifically the independence of residuals which is implied by the skewness-adjusted bootstrap. Previous researchers have documented that marginal decision tends to be cross correlated as well as clustered through time by industry (i.e., post-merge underperformance (Gaspar et al., 2005; Chen et al., 2007) and under-performance of IPOs (Ritter, 1991; and Loughran and Ritter, 2000)). Therefore, they suspect the independence assumption to be violated to some degree. The Jarque–Bera test rejects the normality assumption of the simulated empirical distribution. To understand the severity of this issue, they compute the critical value of normal distribution that is generated from the mean and variance of the empirical distribution. When comparing the critical value of the normal distribution with the critical value of the empirical distribution, the result indicates insignificant difference in the residual variance. However, they carry out a pairwise correlations test of the BHAR to further examine the independence assumption. The result concludes that the pairwise correlations test of the BHAR is relatively small and increases with time across the three marginal decisions samples. However, they warn against the danger of drawing false inferences from ignoring the cross-correlation, specifically with large samples.

Facing the problem of independence, Mitchell and Stafford (2000) advocate the use of CTAR over BHAR to detect long-run performance. They conduct several robustness tests of the CTAR regression to address issues that have been documented in previous research and to ensure consistent estimates of the abnormal performance. To mitigate the heteroskedasticity issue, they impose a minimum number of event firms in the calendar time portfolio to account for diversification effect. Furthermore, Horowitz (1996) bootstrap is deployed to generate the critical value of the t-statistics empirical distribution. The result indicates that the statistical significance of CTAR is unaltered when Horowitz (1996) bootstrap is used.

They also employ dummy variables to account for the number of event firms in each calendar month (i.e., small, or large numbers of firms). They conclude that Loughran and Ritter's concern over false statistical inferences, caused by heavy event activity in certain calendar months, is invalid in their samples. The observed abnormal return is not systematically associated to months with high event activities. Furthermore, to assess the power of the CTAR regression, they induce a constant level of abnormal return to 1,000 random samples of 2,000 firms. The results reveal that the CTAR approach has sufficient power to detect abnormal performance, specifically when the portfolios are formed on value weighted basis.

In short, the CTAR approach alleviates the independence problem inherent in the BHAR approach. However, it is confronted with the constant factor loadings problem, which is imposed by the regression, and it does not reflect investors' true returns.

Liu and Strong (2008) implicitly criticise the CTAR approach and argue that the calendar time portfolio formation method requires rebalancing the portfolio at each point of time. The danger of this formation method is that it produces abnormal return that does not reflect investors' terminal wealth. Furthermore, inferences based on equally weighted calendar time portfolio are flawed or biased. In particular, the result tends to overestimate the premium associated with size factor and underestimate the momentum premium. They strongly recommend that the returns should measure investors' experience and the terminal wealth of their investment. Instead of rebalancing the portfolio periodically, they propose decomposed portfolio returns that maintains the buy and hold characteristic. Specifically, at each point of time the weight of each stock in the portfolio is dependent on the previous holding period performance. Therefore, the portfolio returns represent a passive, buy-and-hold, investment strategy over the holding period.

2.3. Performance Among UK-Equity Unit Trusts

2.3.1. Investment Style Performance of UK Equity Unit Trusts

Mutual funds' investment style is an important approach to describe fund managers' stock selecting behaviour, and to control for the overall risk-return profile of a fund. Typically, there are three broad methods in classifying a fund's style. The first is by simply relying on the fund's self-reported investment style. Brown and Goetzmann, (1997), Elton et al. (2003) and Sensoy (2009) show that a fund's self-reported investment style does not necessarily represents the fund's actual stock holdings. Often, fund managers are engaged in window dressing activities to improve ex-post performance, therefore a fund's self-reported style gives a little information on the actual fund's stock holdings. The second is characteristics-based style analysis, where a fund's style is derived from the characteristics of its stock holdings over a specific period of time. Although this method seems appealing, data on funds' stock holdings are not widely available and are costly (Kaplan,2003). The third is return-based style analysis which decomposes fund return based on the fund's exposures to tradable indices. Typically, a rolling regression is employed to capture the sensitivities of a fund's return relative to a set of passively traded factors. The advantage of this approach is that only the fund return is required to draw a conclusion on the fund's underlying investment style. Therefore, return-based style analysis became very popular not only among academics, but also among institutional investors and advisors.

Due to the popularity of the size and value-growth investment style among mutual fund industry participants, and its tie to the empirical asset pricing literature, many studies have sought to evaluate mutual fund performance using size and value style dimensions. Quigley and Sinquefield (2000) examine the performance of 367 UK-equity unit trusts from 1978 to 1997 by using monthly data obtained from the S&P Micropal database. The average unit trusts return is regressed against the single factor and the three factor-model. The intercept shows that the average UK-equity unit trusts underperform the market index (FTA All Share) by 1.56%/0.48% net/ gross per year using the single factor model, and 2.16%/0.6% net/ gross using the three-factor model. A further investigation is carried out to explore the relationship between unit trust performance and their investment style. Unit trusts are sorted according to their AUTIF (The Association of Unit Trust Investment Funds) sector classification into four equally weighted portfolios, namely, Growth and Income, Growth, Equity Income, and Smaller Companies. The result shows that the three-factor model alpha net return is negative for all AUTIF sector classification. An alternative to AUTIF sector classification, the study uses the three-factor model to capture the sensitivities of unit trust's return to SMB and HML factor. In particular, unit trusts are ranked yearly into 10 decile portfolios by the loading on the size or value-growth factor based on the prior three years returns. The SMB portfolios show that the deterioration in performance increases with the SMB exposure. The underperformance is worsened when unit trusts are tilted toward small stocks, for example unit trusts with the highest prior three-year SMB factor exposure of the three factor-model produces the lowest performance. The HML portfolios show that there is no notable difference across the decile portfolios. They conclude that UK-equity unit trusts have no consistent exposure to either value or growth stock. Finally, they construct nine equally weighted portfolios formed jointly between SMB and HML exposure to capture the interaction effect.

Similar results are found; the underperformance is more pronounced for unit trusts that are tilted toward small stocks.

Davis (2001) investigates the relationship between fund performance and investment style using the Fama and French three-factor model on 4,686 equity funds covering the period 1962 to 1998. Each year, factor loadings are estimated using a 36-month rolling window, then sorted into nine equally weighted portfolios based on the intersections of the SMB and HML rankings. Thereafter, the post-formation return is tested against the three-factor model to identify style-alpha abnormal returns. The results report strong evidence of mutual funds' tendency to favour growth stock over value stocks. For example, HML factor loadings for the top HML decile portfolio is only 0.2. Furthermore, none of the investment style portfolios generate significant alpha, except for the extreme value portfolio. However, in terms of absolute alphas return, growth portfolios performed better than value portfolios across all size classifications.

These findings are also supported by the work of Shi and Seiler (2002), who report that average returns are higher for growth funds than value funds in each of the size classifications. The study investigates the US mutual fund industry with investment objectives of large-value, medium-growth, medium-value, small-growth, and small-value for the period between 1989 and 1999. A mutual fund's style is obtained from the Morningstar Principia Pro database, which classifies funds according to their average stock price to earnings ratio relative to the average of the S&P 500 market index. The average excess return of the six selected investment styles is compared to the Standard and Poor's 500 for large-cap stocks, the Standard and Poor's mid-cap index for mid-cap stocks, and the Russell 2000 for small-cap stocks. The results show that growth funds outperformed value funds by an average of 1.54%. However, growth funds have higher risk than value funds in each of the size classifications.

Chan et al. (2002) provide a comprehensive analysis of the relation between fund's style and performance for 3,336 US equity funds during the period 1989-1997. Both the characteristics-based and returns-based approach are examined based on size, book-to-market, and the momentum characteristics of the underlying funds. Furthermore, in the context of return-based style analysis, the three-factor model and Sharpe's model are compared in terms of the variance difference between the fund's return and the return on its style benchmark. Generally, the result shows that there is ample evidence that the three-factor model and the Sharpe's model give similar identification of a fund's style. The results report that over the universe of funds there is a tendency to cluster around the market index (S&P 500). Although, many funds tend to hold small stocks compared to those in the market index, these funds represent a small percentage of the total market value. The three-and four-factor models show that the intercept alpha of each investment style category is statistically no different from zero. However, growth style funds provide higher absolute performance than value style funds in each of the size classifications. The results suggest that the performance difference between growth and value funds is attributable to past performance. For example, growth style funds buy stocks with good past performance. They further investigate fund's style consistency (correlation) between a fund's past three-year style exposure. There is evidence of style shifting among funds with poor past performance, specifically value-oriented funds. Finally, the returns-based approach performs as well as more elaborate characteristics-based approaches.

However, a characteristics-based approach gives more accurate predictions of future performance when funds' style classifications differ between the two approaches.

Pettengill et al. (2014) compare the performance of growth and value mutual funds in the US using data from 1979 to 2012. Mutual fund information is obtained from the Morningstar database which divides the value and growth funds into three size groups. The data sample is restricted to funds that have a consistent style throughout the sample period. The performance is measured using the total value returns of the Russell 2000 value and growth Index, and Russell 1000 value and growth. Motivated by the fact that investors are interested in their terminal wealth, the geometric and arithmetic mean return is calculated from an initial investment of $10,000 for each style category. Unlike previous US studies, their result shows that value-oriented funds do better than their growth-oriented counterparts, both in terms of relative returns and risk-adjusted returns (Sharpe ratio). Furthermore, the returns gap is related to fund size attribute, therefore investors who wish to exploit the value premium may benefit from an investment in small value-oriented funds. However, the difference in mean returns across the value and growth portfolio is statistically insignificant. The authors also find a direct link between the portfolio's return variability and the geometric and arithmetic mean return. For example, the higher the variability in return the higher the gap between the geometric and arithmetic mean return. They argue that the geometric mean return is a more appropriate estimator to investor terminal wealth. Furthermore, when examining investors' terminal wealth, the study favours total risk (Sharpe ratio) as an alternative to systematic risk measured by factor model. Their argument is that investors are concerned about the variability of the fund's realized return, and factor loadings are meaningless. Additionally, the three-factor model is biased against value funds. For example, the HML factor loading is positive/negative for growth/value funds, and this will lead to decrease/increase expected returns and therefore increase/ decrease the estimated abnormal return.

Brookfield et al. (2015) examine the profitability of UK-equity funds' investment style over the period 1987 to 2010. Forming an equal and value weighted portfolio of the UK-equity funds, the three-factor model reports an aggregate annual alpha of 1.75% for equal-weighted and 1.27% for value weighted funds, gross of fees and transaction cost returns. Using the three-factor model, the study estimates fund's factor exposures over a 52-week rolling window, and then group them into quartiles. The results show that investors are better off holding funds that are more value-oriented than growth-oriented. For example, value-oriented funds outperformed growth-oriented funds by around 0.033 and 0.025 basis points per year for equal and value weighted, respectively. However, when the momentum factor is considered, performance deteriorated across all fund's style category and became statistically insignificant. The result also shows that value and small oriented funds are more likely to buy past winner stocks, and hence to benefit from the momentum premium. To test the fund's style consistency Brookfield et al. (2015) measure the correlation between a fund's current style exposure and its future style exposure. The results reveal inconsistency in fund's styles; there is significant difference in funds' style ranking across the sample period. Similar to Chan et al. (2002), the style shifting is driven by poor past performance.

Overall, the literature overwhelmingly focuses on the investment styles derived from size and book-to-market measures. Tests of the relationship between fund's style and performance have been addressed in a variety of methods in mutual fund literature. Recent empirical evidence suggests that value-oriented funds can earn abnormal returns after accounting for style-risk factors. However, this result may be different between the US and UK literature. For US studies on mutual funds' performance there is little empirical evidence of abnormal performance among value-oriented funds.

2.3.2. Skills VS Luck in Unit Trust Performance

There has been a significant debate over whether fund managers possess superior stock picking skills to generate a higher risk-adjusted return relative to a benchmark. Despite the overwhelming empirical evidence documenting the underperformance of actively managed funds on average, some subgroups of funds do seem to generate abnormal performance relative to their benchmark. However, many researchers have questioned the statistical properties and power of these tests in determining long-run abnormal performance. Empirically, funds' returns exhibit non-normal distribution Warner, (1997) Kothari and Warner (2001) Kosowski et al. (2006) Cuthbertson, et al. (2008). There is no complete model for describing expected returns (Fama, 1998). Furthermore, data snooping may severely bias performance assessment and generate inexistent abnormal performance. Sullivan et al. (1999) define data snooping as a situation where the same set of data is used to construct trading rules and also to test them. The data snooping issue arises when enough trading rules are tested, some of them would generate significant returns, due solely to chance.

Kosowski et al. (2006) propose an advanced statistical bootstrap procedure to evaluate the distribution of fund's return and separate skills from luck in performance. Looking at the net of cost monthly returns of 1,788 US open-end domestic equity mutual funds between 1975 and 2002. The distribution of individual fund residuals is obtained using unconditional and conditional models of performance evaluation. The baseline bootstrap technique is carried out by re-sampling residuals with replacement from the original residual estimates. Then, by imposing the null of zero alpha, the estimated explanatory factors along with the re-sampled residuals are used to simulate a time series of pseudo-monthly excess returns for each fund. This procedure is repeated for each fund 1000 times to generate the cross-section of bootstrapped alphas and t-statistics of alphas. Finally, they compare the p-values generated from the cross-sectional bootstrap with these of individual ranked funds. The results show that, when funds are ranked with the unconditional four-factor model by their alpha (t-statistics of alpha), the top 10% (5%) of mutual funds' performance are significant and cannot be explained by random sampling variation. Furthermore, funds with alphas below the median do significantly underperform their benchmark. However, they favour the t-statistic of alpha over the alpha ranked funds, since it controls for differences in risk-taking across funds. Their study also tests the relation between investment style and skill performance. The results for growth and aggressive growth funds indicate that the top 5% ranked funds are skilful, while funds ranked below the 20th percentile do not have enough skills to cover their transaction cost and management fees. In contrast, there is no evidence of superior stock picking ability in small, income, and all companies' funds. These findings verify previous research in the US market (i.e., Chen et al., 2000) which show superior performance among growth-oriented fund managers.

Cuthbertson et al. (2008) employ the methodology of Kosowski et al. (2006) on a survivor-bias-free sample of 842 UK equity unit trusts/OEICs over the period 1975 to 2002. The average net abnormal performance is negative, but statistically not different from zero. These findings are consistent with previous research that shows that on average mutual funds do not earn enough returns to cover their cost. The result shows significant difference between the parametric (t-test) and non-parametric (bootstrap) test; it is apparent that the difference between the two results is due to the highly non-normal distribution of alphas, particularly at the extreme tails of the distribution. Whereby, the nonparametric approach is based on less restrictive underlying assumptions than the parametric approach. The skilful performance exists only on the top 7 ranked funds, while poor skill performance documented up to the 40th percentile of the performance distribution. The study further examines whether the cross-sectional distribution of alpha changes with fund's investment style. Three investment style categories are considered, namely, equity-income, all companies, and smaller companies. Of the 162 equity income funds, between the highest 3rd and the 10th percentile, funds record a positive performance that cannot be explained by random sampling variation. However, there is barely any skilful performance in funds with all companies and smaller company investment objectives. Furthermore, the majority of poorly performing funds is due to poor skills rather than bad luck.

Fama and French (2010) argue that the baseline bootstrap is biased toward finding skilful performance. The main criticism of the baseline bootstrap is that the independence simulation for each fund result in a loss of the correlation of alpha estimates, specifically when the factor model does not perfectly fit the data. Alternatively, Fama and French propose a bootstrap procedure which resamples residuals alongside the explanatory factors to maintain the correlated movements in alpha estimates. The procedure also eliminates the survivorship bias that may occur because of the minimum fund's observations requirement. Particularly, a fund is included in the data sample if it has 8 months of return observations. To test this bootstrap, Fama and French use a sample data of 5,238 US funds which have existed between the period 1984 to 2006. After controlling for the size factor, the multi-factor model indicates that on net return scheme there is a handful of funds who have enough skills to cover their management fee.

Blake et al. (2015) conduct a comparison between the baseline and Fama and French bootstrap on 561 UK domestic equity unit trusts and OIECS over the period 1998 to 2008. They examine the aggregate performance of equally weighted and value weighted funds with net and growth returns using the four-factor model. They also test for market timing by including the quadratic term of the market excess returns to the four-factor model. Under equal weighted and value weighted schemes, the results report a positive average alpha measure across funds at the gross return level but not net of management fees. However, both net and gross performance are statistically insignificant at any conventional level of significance. The result shows that the additional quadratic term of the market excess returns provides no added value to the results. A comparison of the two bootstrap results reveals that, using growth returns, both bootstraps predict that the performance of the bottom 5th percentile of the alphas distributions result from poor skill rather than bad luck. Similarly, funds that are located above the 70th percentile outperform their luck. At the net return level, although the t-statistic of alphas (parent distribution) have shifted significantly to the left, there is little difference to the t-statistic of alphas distribution produced by either the baseline bootstrap or Fama and French bootstrap. Fund managers stock picking skill completely disappear under the Fama and French bootstrap. While funds that are ranked above the 95th percentile appear to have significant skill under the baseline bootstrap. These findings that skilful performance is more likely to appear using the baseline bootstrap is largely consistent with the findings of Fama and French (2010).

Blake et al. (2015) carry out a further assessment to compare the baseline bootstrap or Fama and French bootstrap. Following Davidson and Duclos' (2000) stochastic dominance technique, they compare the performance distribution of the two bootstraps with each other and with the parent distribution. The results show that the baseline bootstrap is indeed more dominant than the Fama and French bootstrap. For example, the baseline bootstrap is more likely to conclude that the performance is due to good or bad skill than the Fama and French bootstrap. However, when comparing the two bootstrap distributions with the parent distribution, the parent distribution appears to stochastically dominate both bootstrap distributions. Blake et al. conclude that fund managers stock picking skill does exist, but that their performance is wasted by charging higher operating and management fees.

More recently, Huang et al. (2020) estimate US fund performance using the unconditional three and four- factor model and the conditional model of Ferson and Schadt (1996) over the period 1984 to 2006. The data set follows the same period of time that is used in Fama and French bootstrap procedure. Results from estimating aggregate unconditional regressions of equally weighted and value weighted portfolios on a net and gross returns basis are closely aligned with the estimates of Fama and French. Similarly, inferences about the simulated cross-section of true α reveal that the t-statistic of alphas distribution at several points and percentiles are the same as those observed by Fama and French. Under the conditional model, the estimate of the cross-section distribution of t-statistic of alphas has a larger magnitude than the unconditional model in the right tail of the distribution but insignificant effect on the simulated t-statistic of alphas. They conclude that unlike Fama and Frech findings, funds in the right tail of the distribution have genuine stock picking skill and their performance cannot be due to good luck. Thus, restricting the relationship between factor coefficients and the lagged public information have a significant effect on estimation of percentiles of the cross section of the performance distribution.

Song (2020) documents that mutual fund investors allocate capital among funds based on prior factor-related returns. Thus, funds with factor-related returns would lead to larger assets under management and negative fund future performance. Song (2020) conclude that the negative average performance of active funds compared to passive funds documented in Fama and French (2010), are driven mainly by funds with large asset under management that have significantly positive prior factor-related returns.The general view in the literature is that the bootstrap procedure not only assesses making statistical inferences of fund's performance, but also helps in separating out skill from luck in performance. The UK-equity managed funds on average do not earn significant abnormal returns after accounting for operating expenses and management fees. There is strong evidence that poor performance is due to genuinely unskilled managers as opposed to bad luck. However, there are few fund managers with enough skill to generate risk-adjusted performance that cover their costs.

2.3.3. Ethical Unit Trusts

Recently, there has been a growing interest in ethical investing as a result of increased awareness of environmental issues. Corporate scandals, and stock market crashes that led to billions of pounds in losses have further contributed to the growth of this trend. Figures from the EIRIS foundation show that investment in UK green and ethical funds has risen from £4.5bn in 2008 to £23bn in 2019, as environmentally and socially conscious investors continue to look for sustainable and responsible investments (EIRIS, 2021). To cater for the growing demand for ethical investing, many researchers have investigated the performance of ethical funds and compared them with that of conventional funds. Although there is evidence of ethical investors' willingness to sacrifice some return for their noble cause, ethical funds must appeal to investors with different attitudes towards principle and financial reward in order to support sustainable growth in the industry (Berry and Yeung, 2013).

In accordance with modern portfolio theory, imposing ethical constraints on an investment portfolio will limit the construction of the optimal portfolio. As the universe of investments is reduced, ethical fund managers will find it difficult to outperform fully diversified unconstrained portfolios. In addition, the costs of monitoring the activities of the companies they invest in might also cause a further burden to ethical funds' performance. To the contrary, many have argued that ethical funds will benefit from improved economic performance over the long run. The core argument is that high levels of corporate social responsibility are indicators of high quality of management, and this reflects comparative advantages over less socially responsible firms Gregory et al. (1997).Hamilton et al. (1993) suggest three propositions with regards the expected returns of ethical and non-ethical funds. First, both ethical and non-ethical funds have an equal expected return. Expected return varies if and only if the risk factors vary, hence characteristics of a socially responsible company are not priced. Second, ethical funds have lower expected returns than non-ethical funds. In this scenario, ethical funds managers overvalue socially responsible companies and drive down their expected returns compared to non-socially responsible companies. Furthermore, ethical funds are not diversified enough since they are constructed from a subset of the market portfolio. Finally, the expected returns of ethical funds are higher than those of equivalent non-ethical funds. This situation occurs when non-ethical funds' managers underestimate the likelihood of negative information about non-ethical companies as a result of bad social practice.

Most previous empirical works on ethical funds have largely focused on comparing the performance of ethical funds to that of conventional funds. Mallin et al. (1995) examine the performance of UK ethical funds between 1986 and 1993. Based on fund size and inception date, they compare the monthly returns of 29 ethical funds to 29 non-ethical funds. The Jensen's alpha, Sharpe, and Treynor measures are used to understand whether funds which apply ethical criteria are more profitable than non-ethical funds. When the individual funds are adjusted for risk in a single-factor model, both ethical and non-ethical funds underperform the FTASA (Financial Times All Share Actuaries Index). However, when comparing the ethical and non-ethical matched pairs, two third of ethical funds have a higher alpha than non-ethical funds. Similarly, Sharpe, and Treynor measures show that ethical funds are superior to non-ethical funds. Gregory et al. (1997) investigate 18 UK ethical funds' performance between 1986 and 1994 by replicating the methodology of Mallin et al. (1995). The result shows that there is no significant difference in performance between ethical and non-ethical funds.

Similarly, Bauer et al. (2005) report the performance of 103 German, UK, and US ethical mutual funds for the period from 1990 to 2001. Each ethical fund is matched with an equal weighted portfolio of three conventional funds. The sample is matched on the basis of fund age and size as well as their investment focus (domestic vs international). The abnormal performance is estimated from the single and four-factor models, using both domestic and international market indices alongside several ethical indices to compare the explanatory power of these indices. The results show that alphas are statistically insignificant for the three countries and conclude that imposing ethical constraints on mutual funds will not come at the cost of poorer performance. The authors note that ethical funds returns are tilted toward growth companies. Furthermore, unlike the US, ethical funds in the UK and Germany are heavily exposed to small market cap firms. This result is consistent with Guerard (1997) findings that ethical funds managers are biased toward growth stocks. This is because ethical screening criteria are more likely to exclude value stocks, such as chemical, energy, and military equipment. Bauer et al. (2005) further investigates the management fee effect on funds' performance. Although ethical funds charge higher fees than non-ethical funds, there is no significant difference between ethical and non-ethical fund across the three countries.

Gregory and Whittaker (2007) report the performance and persistence in the performance of UK ethical funds. The FT Unit Trust Yearbook is used to collect information of 32 UK ethical funds that have existed at any point between 1989 to 2002. To avoid survivorship bias inherent in a matched pairs approach, each ethical fund is matched with a portfolio of 5 non-ethical funds. Matching criteria include that the matched fund must have a similar inception date and be drawn randomly from the same AUTIF category. The abnormal performance and performance differences are then estimated from the four-factor model, and Treynor and Mazuy approach to investigate fund managers' market timing skill. The results show that both ethical and non-ethical funds underperform their benchmarks. However, there are no differences in alphas between ethical and non-ethical funds under either of the models. Not surprising, ethical funds have a positive exposure to the size factor, while non-ethical funds have statistically significant positive exposure to the HML factor. This is largely consistent with the finding by Bauer et al. (2005) that with regards to performance investors should be indifferent between ethical and non-ethical funds. Gregory and Whittaker (2007) further investigate the persistence in performance using a past performance ranking producer and contingency table approach. The results show evidence of superior performance persistence among the top quartile of ethical funds while failing to find evidence of persistence among non-ethical funds. Blake and Timmerman (1998) propose a possible explanation for the persistence in performance is that ethical funds are tilted toward small companies which generally appear to exhibit predictive power in reference past performance.

An important contribution to the literature on UK ethical fund performance is that of Ferruz et al. (2010). The study investigates the relationship between ethical criteria and financial performance to determine whether ethical pension funds pay higher prices for their ethical consideration compared to non-ethical funds. The study analyses the monthly performance of 28 to 40 UK ethical pension funds that have existed between 2001 and 2007. The alphas are estimated from the four-factor of Treynor and Mazuy (TM) model, and the Merton and Henriksson (MH) model to investigate fund managers' stock picking and timing skill. The market benchmarks used are FTSE4Good Europe for ethical funds and MSCI Europe for conventional funds. Using both models, the authors report an aggregate conventional alpha performance measure close to zero and insignificant. However, the average conventional pension fund managers exhibit negative and statistically significant market timing skill. Similar results are also found for the ethical pension funds. By comparing ethical and non-ethical pension funds, the authors conclude that ethical fund managers have slightly better stock picking ability than non-ethical funds. However, both ethical and non-ethical funds exhibit negative market timing skill.

In conclusion, by comparing the results of UK ethical funds and non-ethical funds studies, one would conclude that a fund's ethical criteria do not seem to cause additional financial costs or benefits compared with non-ethical funds. With regards to financial performance, investors should be indifferent between investing in ethical or non-ethical funds. However, there are significant differences in factor exposures across ethical and non-ethical funds. Perhaps, the most agreed upon factor exposure was that ethical funds are heavily exposed to small-cap firms. Therefore, pair or portfolio matching procedure of ethical and conventional funds must adjust for small companies' exposure in the matching criteria.

2.4. Conclusion

This chapter is dedicated to a review of the literature related to the performance measurement of the UK-equity unit trusts. The chapter outlines the most common performance measurement metrics as well as empirical findings of the most relevant research. Each metric has its own flaws and strengths, offering different explanatory power and highlights a specific aspect of fund performance. Much of the extant literature often reaches conflicting conclusions on fund performance and manager skill. Therefore, the evaluation of fund performance is an extremely difficult matter. The chapter also highlights the most important econometric issues related to long-run event studies. Overall, there is no consensus in the event studies literature on which approach is superior. Often, BHAR and CTAR are viewed as complementary rather than alternatives for robust statistical properties and improved inferences of long-run abnormal return. The next chapter provides some insight into the UK unit trust industry and in particular the data set used in this thesis.

Chapter 3

Data

3.0. Introduction

This chapter provides a detailed description of the data employed in this thesis as well as data definitions and sources. We also provide a simple descriptive statistic of the unit trust returns, and benchmark factor portfolios.

3.1. Overview of the UK asset management industry and investment environment

The UK is the largest investment management centre in European, and one of the top destinations for foreign direct investment. IA members' UK-managed assets grew from just over £2 trillion in 2003 to £9.4 trillion at the end of 2020. Of that, almost 17% (1.6 trillion) of total assets were managed by UK authorised funds. Despite the UK's departure from the European Union, in the last two years, the UK authorised funds industry has experienced an average annualised growth in assets under management of approximately 14.5%. The UK authorised funds industry is highly international, with 44% of total assets are managed on behalf of overseas clients, highlighting the importance of the UK as one of the important international finance centres (theai, 2021). While advantages such as time zone, language, and a stable legal system have helped the UK position as an investment hub for overseas investors. The attractiveness of this industry is based on a regulatory framework that is customer focused, which gives the UK competitive advantages and make it an attractive investment hub for overseas investors. The Investment Management association (IMA) suggests that, such international opportunities are arising thanks in large part to the Financial Services Compensation Scheme (FSCS), competitive tax regime, and technological transformation.

The UK authorised funds industry contributes to efficient market operation by pricing information correctly and bringing transparency to the marketplace. It has also an important role in undertaking stewardship activity over the companies they invest in to protect the value for their clients. In addition, the industry contributes £5-7 billion in tax revenue and create over 114000 job across the sector.

The UK authorised funds industry are structured as unit trusts or open-ended investment companies (OEICs). Unit Trust and OEICs offer investors access to a diversified portfolio of marketable securities that is managed by professional management service. The operation of unit trust scheme is governed by a trust deed with trustees (typically a bank or insurance company) and the manager of the scheme who will be responsible for investing the assets of the unit trust in accordance with the terms of the trust deed. The investors are the beneficial owners of the trust property. Unit trusts are "open-ended" because the managers can "create" or "cancel" units. The value of the unit goes up and down in line with the underlying assets held by the fund. The price of unit is calculated once a day and manager has to choose between dual pricing, single with dilution levy and swinging single pricing. On the other hand, OEICs are governed by company law rather than trust law. OEICs are usually single priced, however dual pricing is permitted too. In recent years, many unit trusts managers have converted to OEICs. The motivation for the conversion is mainly that it costs fund managers less to run an OEIC than a Unit Trust. Furthermore, OEIC's structure allows investors to switch between different OEICs at lower cost than unit trusts.

3.2. Data Sample

The studies of UK unit trust performance are hindered by the lack of reliable return data. In the past few decades only a handful of studies were carried out on the UK funds industry. A potential reason for this might be the survivorship bias issues in the UK data. Most often, commercial databases drop out certain fund's information once the unit trust have exited the market. Brown et al. (1992) showed that when a researcher's sample selection is restricted to funds that survive for the whole period of the study, the performance analysis generates an overly optimistic returns outlook. In the UK, Blake and Timmermann (1998) estimated the average survivorship bias in a sample of 1402 surviving and 973 dead funds.[1] Using an equally weighted portfolio, their results revealed that dead funds tend to have lower mean returns than their surviving counterparts by around 0.8%. Thus, the inclusion of the complete set of funds, both surviving and dead, prevents a possible upward bias in the results when assessing funds' performance.

In the literature, several data sources have been used to evaluate UK unit trust performance, namely, DataStream, S&P Micropal, Morningstar, Money Management Magazine and Unit Trust Yearbook. Table A3.1, shown in the appendix, presents a summary of data sources and details of UK studies of unit trust performance. Our data sample was obtained by combining information from several data sources, namely, DataStream, Trustnet, and the 2002 Unit Trusts & OEIC's yearbook. The list of the

---

[1] A dead fund is one which has existed for some time during the sample period but has not survived until the end of the sample period.

unit trusts currently in existence are obtained from the Trustnet commercial database. The Trustnet database offers information related to fund's performance, size (current), inception date, the Investment Management Association's sector classification, and management fees (not adequate). Our analysis is restricted to the UK All Companies, Equity Income, and Smaller Companies sectors. According to the classification system set out by the IMA, the UK All Companies sector contains funds which invest at least 80% of their assets in UK equities with the objective of achieving capital growth. The Equity Income sector contains funds which invest at least 80% of their assets in UK equity with the aim of having a dividend yield in excess of 110% of the yield of the FTSE ALL Share index. Funds that invest at least 80% of their assets in the smallest 10% by market capitalisation of UK equities are considered to be in the Smaller Companies sector under the IMA classification. Thus, an accurate benchmark portfolio can be constructed to evaluate fund's risk-adjusted returns.

Data on dead unit trusts and OEICs is unavailable on the Trustnet commercial database, meaning that our analysis would be biased by selection towards the current opportunity set of investments. Therefore, we relied on the 2002 Unit Trust & OEIC's yearbook[2] to select those funds that existed in 2002 with UK All Companies, Equity Income, and Smaller Companies' investment objective. Nonetheless, survivorship issues remain since funds that were incepted after 2002 but defunct before the end of the study period were ignored. DataStream offers raw data of 8700 UK dead funds that

[2] This is the official yearbook of the Association of Unit Trusts and Investment Funds (Presently called IMA) and contains a list of unit trusts' SEDOL code, fund size (in 2002), IA classification, investment style and management fees.

existed in their database. Accordingly, a complete population of dead funds over the period between 2002 and 2017 were manually collected and filtered according to fund name and inception date. For example, the study dropped all dead funds whose names contain any of the following words, international, Euro, global, Asia, Pacific, balanced, fixed income, bond, tracker, index, pension, life, insurance, and assurance.

Mutual fund liquidation (dead funds) involves the sale of all of a fund's assets and the distribution of the proceeds to investors. There are a variety of reasons for fund liquidation, with poor performance ranking as one of the primary causes. If poor performance is a cause of liquidation, then conditioning findings on live funds may induce an upward bias in average performance (survivorship bias). Careful examination of our data sample reveals that the return observations of dead funds in last few months of trading show a declining return pattern. This suggests that, on average, funds are liquidated due to poor performance. Blake and Timmermann (1998) show that of the funds that died over the period 1972- 1995, almost 90% were merged with other funds and only 10% were closed. This route is easier for investors because their money is immediately invested in a more successful fund.

Since our goal is to investigate fund managers' stock picking ability, index tracker, pension, and life insurance funds were also excluded. These funds are subject to specific commitments and regulations that govern their permitted investments and asset holding. Furthermore, most funds have multiple share classes. Typically, fund companies issue different share classes for funds to offer a wider choice to both institutional investors and retail investors, in terms of dividend pay out and cost structures. However, fund manager's investment decisions are not independent across different asset classes. Thus, to avoid duplicate counting, the study identified the primary share class of each fund then removed non-primary share classes from the sample, therefore eliminating bias towards funds with more asset classes in equally weighted portfolio returns. The return of funds that have changed name or merged are treated as a continuation of the original trust. Dead, delisted, and suspended unit trusts are treated as liquidation, which involves the sale of all fund assets and the distribution of these proceeds to shareholders.

Finally, the study cross-checked the list of funds from DataStream, Trustnet, and the 2002 Unit Trusts & OEIC's yearbook using SEDOL code, in order to avoid duplicate counting. Thus, our data sample is free of survivorship bias as it includes both surviving and dead funds. Similar procedure to those described above are applied to obtain data on UK equity Ethical funds. A list of ethical funds that have exited in the UKwere obtained from the EcoMarket database. The screening process resulted in a total sample of 352 unit trusts and OEICs, over the sample period January 2002 to July 2017. The sample contained 320 conventional funds comprising of 132 dead funds, 116 UK ALL Companies, 17 equity income and 55 Smaller Companies. Of the 32 ethical funds, 12 were dead, 16 were UK All Companies, and four were Equity Income.[3]

Fama (1976) showed that monthly returns are closer to normality than daily return. Keswani and Stolin (2008) on the other hand compared monthly and quarterly data in capturing fund managers stock selection ability. Their results showed that higher frequency data makes it more difficult to detect managerial skills, beside higher

---

[3] EcoMarket database does not offer data on funds with Smaller Companies investment objective.

observations increase the power of their statistical tests. Consequently, to perform our analysis, the time series data of monthly fund's gross returns were retrieved from the DataStream database. Total return index (RI) is used to calculate the percentage change in monthly individual fund's returns. The total return index (RI) is the theoretical growth in the value of unit holding, assuming that dividends are reinvested to acquire new units at the closing price of the ex-dividend date.

$$RI_t = RI_{t-1} \frac{PI_t}{PI_{t-1}} * (1 + DY_t) \tag{3.1}$$

where PI_t is the bid price index on month t, and DY_t is gross dividend yield of the price index.

The percentage change of unit trust return is therefore calculated as follows:

$$R_{it} = \frac{RI_t - RI_{t-1}}{RI_{t-1}} \tag{3.2}$$

Finally, the monthly data set of market return, risk-free return, as well as factor mimicking portfolios for size, and book-to-market are obtained from Exeter University's Centre for Finance and Investment. The market return is the total return (inclusive of dividends) on the FTSE All Share Index whilst the risk-free return is the return on a one-month Treasury Bill. The FTSE-All Share is the most appropriate and commonly used benchmark by funds in the UK All companies sector, due to its comparability with the fund's returns (Bryant and Taylor, 2012). The FTSE4Good UK index is chosen as the market benchmark for ethical funds. The FTSE4Good is designed to reflect strong economic social and corporate governance (ESG) management practice, whereby companies in the FTSE All-share index are required to score an ESG rating of 3.3 to be added to the FTSE4Good UK index. Factor mimicking portfolios are the Fama and French portfolios of UK stocks sorted on market capitalization, and the book-to-market ratio.

In light of Fama–French NYSE break point, Gregory et al. (2013) use the median of the largest 350 companies as a cut-off point for the size break point. The reason for this is to avoid forming factor portfolios that are being heavily weighted by illiquid and small stocks. Thus, the largest 350 firms are sorted into two groups by market capitalization. The two groups are then sorted by book to market ratio using the 30th and 70th percentiles of the largest 350 firms. They therefore form six intersecting portfolios, namely, small-high (SH), small-medium (SM), small-low (SL), big-high (BH), big-medium (BM) and big-low (BL). These portfolios are rebalanced on an annual basis. Thereafter, using the value weighted principle, the Fama and French's factor mimicking portfolios are calculated as follow: The SMB factor is (SL + SM + SH)/3 – (BL + BM + BH)/3 and the HML factor is (SH + BH)/2 – (SL + BL)/2.

3.3. The Expenses

Investors in UK unit trusts incur two types of charges. The first is paid by investors for professional services provided and are known in advance. This includes various fees such as registration, audit, depositary fees, and an annual management charge, which together make up the total expense ratio (TER).[4] Investors are also subject to initial charges, exit charges and performance fees. However, initial charges are often

[4] This term is being replaced by "on-going charges" under the new Key Investor Information Document.

rebated while exit charges and performance fees are rare. Furthermore, these charges are on the decline since the introduction of the retail distribution review (RDR)[5] (Bryant and Taylor, 2012). The second charge is incurred by investors in order to access the market. This charge includes bid-ask spread, dealing costs and stamp duty. Although the nature of the market being accessed, in terms of liquidity and tax regime, is a key determinant of these costs, the magnitude of the costs varies for different funds depending on the fund manager's level of trading and/or investment style (IMA, 2012).

Data on the first type of charges is published for each fund in the Key Investor Information Document (KIID). While data on the second charges can be accessed from funds annual reports and accounts. Unfortunately, information on these two charges is not easily attainable, particularly for those funds which no longer exist. Bryant and Taylor (2012) investigated the costs figures of the 15 largest FTSE100/FTSE All-Share tracker funds and the 15 largest active funds which represent 86% and 51%, respectively, of the total value of funds invested in the UK All companies' sector. Based on 2011 funds' accounts data, their results showed that the cost of FTSE-All Share (FTSE100) tracker averages 0.69% (0.84%) per year. This consists of 0.62% (0.75%) type one charges and 0.07% (0.9%) type two charges. The cost of active funds averages 1.95% per year of which 1.57% constitutes type one and 0.38% type two charges.

---

[5] The Financial Conduct Authority has introduced RDR in 2012 to remove the incentive for product bias across the retail investment advice sector.

Based on Bryant and Taylor (2012) findings, this study deploys fund's average charges to estimate their net returns. We assume that fund's charges are constant over the sample period from Feb 2002 to Jul 2017. Specifically, each month we calculate the net returns on active/passive managed funds by deducting the corresponding average charges from the fund's raw return.

3.4. Descriptive statistics

In this section, we provide the descriptive statistics of the aggregate returns of the live, dead, and the entire sample of funds over the period from February 2002 to July 2017. The monthly equally weighted fund's returns are measured in excess of the monthly treasury bills and are gross returns.

Table 3.1 shows the summary statistics of the mean monthly excess gross returns, skewness, kurtosis, Jarque-Bera (J-B) test, and the proportion of funds which reject the normality assumption across three equally weighted portfolios. In Panel A, the descriptive statistics reveal that the average monthly gross excess return for all conventional funds is 0.47%. The average monthly return for surviving funds is 0.57% compared to 0.30% for dead funds. This result supports the hypothesis that short lived funds are more likely to exit the market as a result of prior poor performance. The variation of returns is marginally similar across all three aggregate portfolios, with a standard deviation of 3.8% per month. The return distribution exhibits slightly negative skewness, and high excess kurtosis. With regards to the normality question, the normality of return distribution is strongly rejected by the Jarque Bera test. Our result shows that normality is rejected for around 69% of all conventional, 82% of surviving and 51% of dead funds. Kosowski et al. (2006) explained this as being due to funds holding of derivatives to hedge returns, or that the co-skewness of individual non-normal stock returns may not be diversified. Consequently, inferences based on the standard parametric tests may not be reliable in assessing funds' performance. Panel A also shows the moments for the monthly gross excess returns of the market portfolio. The FTSE-All Share mean return is slightly lower than the aggregate for all conventional funds. However, the FTSE-All Share portfolio is a passive buy and hold industry portfolio, and as we showed earlier it offers significant lower charges than active funds. The FTSE-All Share has the highest standard deviation at 4.15% per month. The high standard deviation might suggest that the benchmark portfolio is capable of explaining the variation in funds' returns.

Similarly, Panel B presents description statistics for ethical funds. The average monthly excess return of the whole market accounts for 0.37% per month. The surviving funds have an average monthly return of 0.39% compared to 0.27% of monthly return for average dead funds return. However, the size of survivorship bias is marginally smaller than those observed in Panel A. We also observed similar return variation when comparing the monthly standard deviation across the two panels. Negative skewness and significantly positive excess kurtosis are more pronounced for ethical funds than its conventional counterpart. However, a higher rejection rate of normality for ethical fund's return was to be anticipated, considering that ethical funds are less diversified than their conventional counterparts. The FTSE4Good UK index is considered as the benchmark for ethical funds; FTSE All-Share companies need to meet environmental, social and governance risk management practices to be included in the index. The average monthly return of the FTSE4Good index is 0.39% per month, achieving higher returns than the aggregate ethical funds. However, we also observed higher variation in its returns than those of its ethical counterparts.

Table 3.1: Descriptive statistics of the monthly excess gross returns for the aggregate sample of funds, survivorship, and market portfolio

Panel A: Conventional Funds				
Series	*All funds*	*Live funds*	*Dead funds*	*FTSE-ALLSHARE*
Number of Funds	320	188	132	-
Mean (%)	0.47	0.57	0.3	0.45
Std Error (%)	3.8	3.9	3.8	4.15
Skewness	-0.86	-0.77	-0.91	-0.64
Excess-Kurtosis	1.53	1.48	1.46	1.2
Jarque-Bera P-value	0.00	0.00	0.00	0.00
Rejection of Normality % of funds	69	81.9	50.7	-
Panel B: Ethical Funds				
Series	*All funds*	*Live funds*	*Dead funds*	*FTSE4GOOD*
Number of Funds	32	20	12	-
Mean (%)	0.37	0.39	0.27	0.39
Std Error (%)	3.73	3.92	3.68	4.12
Skewness	-0.97	-0.93	-0.91	-0.56
Excess-Kurtosis	2.14	2.15	1.95	1.1
Jarque-Bera P-value	0.00	0.00	0.00	0.00
Rejection of Normality % of funds	81.25	90	66.7	-

It seems that, on average, fund's managers were able to provide better gross returns than the market. However, both performances are statistically insignificant at any conventional level of significance. The coefficient of excess market return (FTSE-All Share/FTSE4Good) is positive and statistically significant at the 1% level, it explains 86% and 82% of the variation in conventional and ethical funds' returns, respectively. At individual fund level, out of 320 conventional funds, 104 (6) (around 32.5% (1.9%))

funds achieved statistically significant positive (negative) risk-adjusted gross alpha. Also, out of 32 ethical funds, 5 (1) (around 15.6% (3.2%)) ethical funds produce a statistically significant positive (negative) risk-adjusted gross alpha.

The 3-factor model reveals similar findings to those observed under the single factor model. The average gross alpha is positive but statistically insignificant for both conventional and ethical funds. The coefficient of the market factor accounts for 85% and 79% of the variation in conventional and ethical funds returns, respectively. Although positive premium is observed across both SMB and HML factor mimicking portfolios, the standard t-test shows both to be statistically insignificant in terms of explaining the variation in fund returns. By comparing the significance of individual funds abnormal return across the CAPM and the 3-factor model, the 3-factor model produces a similar significance rate to those observed under the CAPM. It indicates that the performance is positive (negative) statistically significant for 30.9% (3.4%) of conventional funds, and 15.6% (3.2%) of ethical funds.

These results are consistent with previous studies in the UK market such as those carried out by Fletcher (1995), Becker et al. (1999), Fletcher and Forbes (2002), and Byrne et al. (2006), who all found no evidence of abnormal returns for active funds. Furthermore, one would expect the performance to be worse once fund charges are considered. Thus, our results might suggest that the UK market is informationally efficient, and that investors are on average better off investing in index trackers. However, at the individual fund level, the performance distribution indicates that some funds do indeed produce statistically significant abnormal returns. To conclude, throughout this thesis, the market model and the three-factor model are applied together to detect risk-adjusted performance and fund manager’s stock picking skills.

Many Studies have showed that the momentum factor mimicking portfolio is statistically insignificant in the UK funds industry (Blake and Timmermann, 1998; Tonks, 2005; Cuthbertson et al., 2008). Accordingly, we decided to ignore the 4-factor model from our analysis. Of course, there remains a joint-hypothesis problem of concerning whether the performance model is the true model of equilibrium. Model misspecification is a likely cause of any cross-fund residual correlation. However, in this thesis many alternative equilibrium models of performance are tested and hence results are unlikely to be model specific. Furthermore, regardless of model specification, the focus in this thesis is on abnormal returns sensitivity to style-adjusted mutual fund returns. Notwithstanding, test statistics derived from the market and three-factor model are not reliable in assessing funds' performance, particularly when returns exhibit non-normal distribution. Hence, several bootstrap approaches will be used to control for non-normality in fund returns and draw reliable inferences on funds' performance.

Table 3.2: Average performance of the aggregate sample of UK-equity Conventional and Ethical funds.

	Market Model	
Coefficients	***Conventional Funds***	***Ethical Funds***
Intercept *t-stat*	0.08% 0.84	0.04% 0.34
β_1 *t-stat*	0.86 26.6*	0.82 19.6*
Adj-R²	88.6%	81.7%
% of funds with significant positive (negative) alpha	32.5% (1.9%)	15.6% (3.12%)
	3-factor Model	
Coefficients	***Conventional Funds***	***Ethical Funds***
Intercept *t-stat*	0.07% 0.80	0.03% 0.27
β_1 *t-stat*	0.85 28.40*	0.79 21.00*
SMB *t-stat*	0.02 0.64	0.07 1.52
HML *t-stat*	0.03 0.64	0.001 0.03
Adj-R²	88.5%	82.3%
% of funds with significant positive (negative) alpha	30.9% (3.4%)	15.6% (3.12%)

The table reports the results from the estimation of the market model and the 3-factor model over the period between Feb 2002 to Jul 2017. The 3-factor regression equation is given as: $R_t - Rf_t = \hat{\alpha} + \hat{\beta}(Rm_t - Rf_t) + \hat{\beta}_{smb}(SMB)_t + \hat{\beta}_{hml}(HML)_t + \hat{e}_t$ where: R_t is the equally weighted portfolio of conventional/ethical funds, Rf_t is the return on one-month Treasury Bill, $\hat{\alpha}$ is the estimated Jensen alpha, Rm_t is the FTSE-ALL Share/FTSE4Good, SMB_t and HML_t are the Fama and French factor mimicking portfolio, and for the single factor model $\hat{\beta}_{smb}$= $\hat{\beta}_{hml}$= 0. * Indicates significance at the 1 percent level.

Chapter 4

Methodology

4.0. Introduction

This chapter presents the methodologies applied in the literature and in this research to evaluate mutual fund performance. The chapter starts with a description of the return-based style analysis approach. The basis of this approach is to compare fund's returns to the returns of a set of asset classes indexes and draw a conclusion on fund managers' investment style. The motive of this approach is to categorise funds based on their style exposure and select an appropriate benchmark for future expected return accordingly.

Section two introduces the event-time methodology as outlined Lyon et al. (1999). The abnormal performance is measured based on the aggregate returns of funds' investment style. Thus, the abnormal performance could be achieved by systematically buying units in funds with specific investment style over the short and long-run. This approach has the advantage of measuring investors' true returns on the underlying investment strategy. It maintains the buy-and-hold property, and therefore properly indicates a typical investor's end wealth from adopting such a strategy. Furthermore, the wild-bootstrapping technique is described, which purports to obtain valid estimates of the abnormal performance.

In section three, the calendar-time methodology is presented, where for each calendar month the abnormal return is calculated as the mean abnormal time-series of event funds' portfolio returns. This approach seeks to eliminate the potential dependence of returns on cross-sectional analysis. The statistical inferences are also robust to heteroscedasticity bias. Specifically, the standard errors are robust using the OLS with White's correction and Gregory et al.'s (2010) Feasible GLS techniques.

Finally, the chapter illustrates the bootstrap methodologies used in separating out luck and skill in funds' performance. Two bootstrap procedures are described, namely, the 'baseline', and the 'skewness-adjusted and kurtosis preserving wild' bootstrap. While the residuals from the performance models are likely to be non-normal, the significance of abnormal performance is tested using a non-parametric approach as opposed to the standard t-test. Thus, the bootstrap approach provides statistical validity to the performance metrics and ascertains whether funds performance is due to skill or luck.

4.1. Investment Style

A large body of mutual fund performance literature has examined the possibility of identifying fund managers who can generate positive abnormal performance in the future based upon their past returns. One popular way to achieve this is to decompose fund's sources of returns over time to infer fund's investment style. In this approach, a multivariate regression can be used to identify risk factors or investment styles that may influence future fund performance. Since the relationship between funds attribute and performance may vary over time a rolling regression is typically employed to produce time varying coefficients. Thereafter, funds' returns are sorted in portfolios based on key characteristics then tested for a successful ex-ante investment style.

The usefulness of this approach is that it helps to determine a fund's risk-return profile, and also helps in evaluating managers' stock selection skill. For example, if there is enough commonality between the fund's return and a certain investment style, then the difference in future performance relative to that predicted from the style adopted is representative of the manager's stock selection skill.

4.1.1 The Sharpe Model

The Sharpe (1992) asset-classes factor model is used to capture fund exposure to variation in returns of a variety of different equity classes. This method is commonly used and accepted in the investment community and forms the basis of Return Based Style Analysis (RBSA) (Chan et al., 2002). In this approach, the majority of fund managers are restricted to invest in predefined asset classes and the only discretion allowed is to select windfall shares within each asset class. Thus, unit trusts' returns are expected to be highly correlated with the returns of standard asset classes. The scope of this thesis concerns domestic UK equity funds. Fund managers' overall asset allocation can be compared with four equity classes: (i) portfolios containing small-cap growth stocks (SG); (ii) portfolios containing large-cap growth stocks (BG); (iii) portfolios containing small-cap value stocks (SV); and (iv) portfolios containing large-cap value stocks (BV). These mimicking indices are widely implemented by pension plan sponsors as a summary measure of pension risk in screening fund managers (Bassett and Chen, 2001). Furthermore, index providers such as Standard & Poor's, Russell, MSCI, and Morningstar have all calculated fund style indices in a way that reflects the size and value-growth orientation of the underlying shares. Most US RBSA studies have replicated the performance of a managed portfolio by the return on Russell 1000 and Russell 2000 indices. For example, Sharpe (1992), Chan et al (2002), and Ben Dor et al. (2003) have all used Russell indexes in their study. The Russell 1000 index is a value-weighted index of the largest 1000 U.S. stocks, and the Russell 2000 index is a value-weighted index of the next 2000 largest U.S. stocks. Since equivalent style benchmark indices are not available in the UK, we used factor mimicking portfolios of size, and value-growth orientation obtained from Exeter University's Centre for Finance and Investment. As explained in the data chapter, the largest 350 firms are sorted in two groups, according to their market capitalization.[6] Each group is then sorted by book to market ratio to form six intersecting portfolios, namely, Small-High (SH), Small-Medium (SM), Small Low (SL), Big-High (BH), Big-Medium (BM), and Big-Low (BL). These portfolios are formed using a value weighted scheme and are rebalanced on an annual basis. Accordingly, we chose four mimicking portfolios to reflect the returns on size and value-growth style indices, SG, SV, BG, and BV: where "SG" denotes small-growth and return is identical to SH, "SV" denotes small-value and return is identical to SL, and "BG", and "BV" denotes big-growth, big-value and return is identical to BH and BL, respectively.[7] All four style mimicking indices are aligned with Sharpe's asset-classes index selection criteria. First, the four style mimicking indices cover most of the investment universe available to fund managers. Second, they are mutually exclusive, for example, no stock is represented in more than one index, and

[6]Gregory et al. (2013) use the median firm in the largest 350 companies (excluding Financials) as a cut-off point for the size and book-to-market break point.

[7] I refer the reader to Gregory et al. (2013) for more details on portfolios construction.

investors can replicate these indices. Assuming that the four factors, SG, SV, BG and BV indices closely explain the return behaviour of the funds being analysed, the Sharpe's asset-classes factor model is obtained by regressing a fund returns R_{it} on the four factors as follows:

$$R_{it} = \alpha + \omega_{i,s}^{SG}\, SG_t + \omega_{i,s}^{SV}\, SV_t + \omega_{i,s}^{BG}\, BG_t + \omega_{i,s}^{BV}\, BV_t + e_{it} \tag{4.1}$$

$t = s - 36, \dots, s - 1\,,\ i = 1,2, \dots, N,\ \ s = 37, 49, 61 \dots, 157$

where R_{it} is the return on fund i in period t, α is a measure of risk adjusted or abnormal fund return, SG_t is the return on small-growth mimicking portfolio, SV_t is the return on small-value mimicking portfolio, BV_t is the return on big-value mimicking portfolio, and BG_t is the return on big-growth mimicking portfolio. Given $j = SG, SV, BG, BV$ the $\omega_{i,s}^{j}$ coefficients represent the sensitivity of R_{it} to factor *f* mimicking portfolio in year s. For example:

$$\omega_{it}^{SG} = \left. \partial E(R_{it}) \middle/ \partial(SG_t) \right.$$

Thus, $\omega_{i,t}^{SG}$ quantify the impact of SG style exposure on the expected return of fund i at time t . The residuals e_{it} represent the difference between fund's return and the return on style mimicking portfolios. The residuals have been termed as the non-factor returns and measure fund managers' stock selection or market timing skills. Most importantly, the non-factor return for one fund is assumed to be uncorrelated with that of other funds. Hence, the only source of return correlation is the asset-classes mimicking portfolios (Sharpe, 1992).

Sharpe (1992) proposed the following restrictions on the regression coefficients:

$$\sum_{j=1}^{k} \omega_{i,s}^{j} = 1 \text{ and } 0 \leq \omega_{i,s}^{j} \leq 1 \tag{4.2}$$

where $j = SG, SV, BG, BV$.

The regression coefficients are constrained to non-negative values and sum to one. Thus, the coefficients not only signify the exposure to different asset-classes but also can be interpreted as fund's weights. In order to obtain a return similar to the fund being analysed, one would invest in a portfolio with $\omega_{i,s}^{j}$ proportion of style j. Ter Horst et al. (2004) showed that the non-negativity constraint on the factor loadings produces more efficient parameter estimates. In effect, short selling is limited in Unit Trusts and the OEICs industry and must be reported with an indication of its use (FCA, 2019). Almazan et al. (2004) report that about 70% of mutual funds of U.S. domestic equity funds from 1994 to 2000 are discouraged from pursuing any short selling activities and only 2% are involved in such activities. The factor loadings are restricted to sum to one, so they can be scaled as weights in analogue portfolios. Thus, fund's return cannot be exposed to a specific style more than the mimicking portfolio index. Ter Horst et al. (2004) recommended the use of both constraints if the purpose of style analysis is to identify the best mimicking benchmark.

Sharpe (1992) applied this procedure in two ways; first he assumed constant weight over the time period covered. Second, a monthly rolling window was employed in order to investigate the behaviour of a manager's average exposures to asset classes over time. For example $\omega_{i,s}^{j}$ represents the average contribution of style j to fund i for month s. Thus, the monthly deviation of a fund's return from its asset-classes mimicking indices is attributed to fund manager's stock-selection of specific shares within the asset-class and/or their market timing ability; for example by rotating across different investment styles.

Since the standard regression analysis is inappropriate when constraints are imposed, Sharpe (1992) applied a quadratic programming technique to estimate fund average exposures to asset-classes in the presence of inequality constraints in equation (4.2). Thus, $\omega_{i,s}^{j}$ are estimated in a way that minimizes the unexplained variation in fund returns e_{it} under the stated restriction. Nonetheless, under quadratic optimisation, the confidence intervals of factor's weight are not readily available. Lobosco and diBartolomeo (1997) used Taylor expansion to examine the statistical accuracy "confidence intervals" of the estimated factor loadings. However, the asymptotic distribution is invalid if the true style exposure is zero or one, in which case the actual factor loadings are on the boundary of the Taylor expansion parameter space.

In our approach, nonlinear optimization is used to explain the behaviour of fund's return by minimizing the variance of the non-factor return e_{it}. Alternatively, $\omega_{i,s}^{j}$ have been calculated in a way that minimizes the fund tracking error "active management effect" over the style benchmarks. Hence, the factor loadings are estimated by the maximum likelihood (ML) procedure using the BFGS (Broyden, Fletcher, Goldfarb, and Shannon) algorithm. BFGS algorithms belong to quasi-Newton methods, while the BFGS optimisation method determines the first and second derivatives of the log-likelihood function with respect to the parameter values at each iteration, known as the gradient and Hessian matrix. Although the

BFGS algorithm is one of the most popular Quasi-Newton or even second-order optimization algorithms used for numerical optimization, the reliability of this parameters' estimation method is questionable. One way to judge the BFGS' ability to capture fund styles weight is to look at the dispersion across funds' tracking error volatilities. By comparing mutual fund returns against the estimated implicit benchmark and against a general benchmark such as the FTSE 350, one would draw a conclusion on whether BFGS method produces a reliable estimation of fund style weights. Accordingly, the estimated style benchmark and the FTSE 350 are selected in this thesis as the `benchmark' model. The BFGS initial guess of fund exposures is set at $\omega_{i,s}^{SG} = \omega_{i,s}^{SV} = \omega_{i,s}^{BG} = \omega_{i,s}^{BV} = 0.25$, which corresponds to the case where fund i is exposed equally to all four mimicking portfolios. Under the assumption of normality, the ML estimated factor loadings are consistent, and the asymptotic distribution can be used to identify fund exposure to a particular style.

We allow factor weights to be time-varying using 36-month rolling windows. Factor exposure to the mimicking portfolios is estimated over a 3-years period then tested for the subsequent 12-month. Thus, unit trusts with return information of less than 48 months are excluded. In other words, a fund was only included in the data set if it had at least 4-years of observations, hence factor exposure across the four mimicking portfolios can be estimated over a 3-year' period then tested for the subsequent 12 months.

4.1.2. Continuous Changing Style

At each calendar year from 2005 to 2017, a customized benchmark is constructed for each fund using the previous 36-month returns. Hence, the mimicking portfolio weight is allowed to vary on a yearly basis but held constant within the year.

Next, we construct portfolios of unit trusts via two approaches. In the first approach, we divide funds into investment categories based on the highest factor exposure produced by RBSA regression. For example, if the highest loading in a particular regression is $\omega_{i,s}^{SG}$ then the returns of fund i are allocated to portfolio C4SG for the 12 months following year corresponding to s. In this approach we have a maximum of four style equally weighted portfolios (C4SG, C4SV, C4BG, and C4BV). We call this approach C4.

In the second approach, we group funds according to the following rule:

(i) If the highest loading is greater than 0.50, we take it as representative of one of the four main styles (SG, SV, BG, and BV).

(ii) If the highest weight is less than 0.50, we take the highest two weights.

This approach leads to a maximum of nine portfolios (shown in Table 4.1). We call this approach C9. For example, if the RBSA regression result for fund i showed that $\omega_{i,t}^{bg}$ is greater than 50%, then the fund's investment style is described as a pure big-growth investment style and the monthly fund's return for the 12 months following year s is included in the C9BG stylized-portfolio. If the RBSA regression result for fund i showed that no factor exposure is greater than 50% and the highest two loadings are $\omega_{i,t}^{bg}$ and $\omega_{i,t}^{bv}$ then the fund's investment style is described as the big investment style and the subsequent 12 months fund's returns are included in the C9B stylized-portfolio.

Table 4.1: Describe the possible investment style according to C9 approach.

Highest possible ranking (Stylized portfolio)	**Fund's investment style (benchmark" mimicking portfolio" obtained from Exeter University's Centre for Finance and Investment.)**
small-growth (C9SG)	small- growth orientation (SG)
small- value (C9SV)	small-value orientation (SV)
big- growth (C9BV)	big- growth orientation (BG)
big- value (C9BG)	big-value orientation (BV)
small-value & small- growth (C9S)	small stocks orientation $S_t = (SG_t + SV_t)/2$
big-value & big- growth (C9B)	big stocks orientation $B_t = (BG_t + BV_t)/2$
small-value & big-value (C9V)	value stocks orientation $V_t = (SV_t + BV_t)/2$
small-growth & big-growth (C9G)	growth stocks orientation $G_t = (SG_t + BG_t)/2$
small-value & big-growth or big-value & small-growth (C9MIX)	Mix investment orientation $Mix_t = ftse\ 350$

One problem encountered in the portfolio formations is that, for some years, some funds cannot be allocated to specific investment style classifications. For example, there is no fund tilted toward growth stocks in 2005. However, these 'absences' are not a result of fund bankruptcies but rather of investment styles shifting. Thus, to compute stylized portfolio returns, we replaced the returns of that year by the risk-free rate. In other words, we assume that investors in that particular style would earn the risk-free rate. Liu and Strong (2008) replace delisted firm return by either zero or the risk-free rate. The result shows no significant difference between the two approaches. In contrast, Mitchell and Stafford (2000) replace all de-listed firms by the benchmark return. However, we argue that this approach has the potential to create an upward bias in the estimated abnormal returns, since poor performance is the primary causes of most dead funds.

For example, the first three years returns of a typical Small-growth portfolio, $C9_t^{SG}$, would look like

$$r_{s+1}^{(5)}, \dots, r_{s+12}^{(5)}, r_{s+13}^{f}, \dots, r_{s+24}^{f}, r_{s+25}^{(8)}, \dots, r_{s+36}^{(8)} \dots.$$

The above example reflects the following. In the first year, 5 funds were identified as Small-growth and the equally weighted average return ($r_t^{(5)}$) was used for the first 12 months. In the second year, no funds were identified as small-growth and the 12 months returns were the risk-free rates (r_t^{f}). In the third year, 8 funds were identified as small-growth and the average return of the 8 funds were inserted in the $C9_t^{SG}$ portfolio return.

Finally, unit trusts relative performance is evaluated by comparing the monthly average stylized-portfolios returns with their corresponding investments style mimicking portfolios.[8] For example, for small-growth style, the abnormal return, AR_t^{SG}, is given by

[8] A discussion around cumulative abnormal return and buy and hold abnormal return will be presented in Chapter 6 and 7.

$$AR_t^{SG} = C9_t^{SG} - SG_t \tag{4.3}$$

Similarly, the abnormal return, AR_t^{Mix} is given by $AR_t^{Mix} = C9_t^{Mix} - Ftse350_t$. We also reported the index based abnormal return, which is obtained by subtracting a given index return from the stylized portfolio returns. For example, the index (FT100) based abnormal return, IAR_t^{SG}, is given by

$$IAR_t^{SG} = C9_t^{SG} - FT100_t \tag{4.4}$$

These differences allow us to attribute performance to investment style; it represents the active management effect or fund's average selection return. Assuming the mimicking portfolios are an efficient style portfolio, then excess return generated over/under the style mimicking portfolio is attributed to funds' managers selection skill/incompetence. Nonetheless, a positive average abnormal return does not necessarily indicate that it is optimal for investors to invest in the stylised portfolio if the choice is limited to invest in either the stylized or the mimicking portfolio. This is because stylized portfolios contain residual risk relative to the mimicking portfolio, which may or may not be correlated with the factor returns. Since the mimicking portfolios are considered as the best benchmarks for stylized portfolio, the stylized portfolios returns are compared to the mimicking portfolios as follow (using SG as an example):

$$C9_t^{SG} - R_t^f = \hat{\alpha} + \hat{\beta}(SG_t - R_t^f) + \hat{e}_t \tag{4.5}$$

where $\hat{\alpha}$ is the expected excess return of the stylized portfolio relative to the mimicking portfolio SG_t. If $\hat{\beta} = 1$ this implies that $var[\hat{e}_t]$ will tend to be small and the mimicking portfolios closely tracks their corresponding stylized portfolio. Hence, a positive $\hat{\alpha}$ indicate that investors with this particular style are better off holding the stylized portfolio than the mimicking portfolio. However, a positive $\hat{\alpha}$ does not necessarily mean a superior performance because the model may not appropriately reflect systematic risk. R-squared is the proportion of variance in stylized portfolio explained by the mimicking portfolio, while 100%-R-squared represents how active the fund managers are.

Summary: Continuous Changing Style

For each fund $i = 1,2, ..., N$, find the weights $\omega_{i,s}^{j}$ from the regression

$R_{it} = \omega_{i,s}^{SG}\, SG_t + \omega_{i,s}^{SV}\, SV_t + \omega_{i,s}^{BG}\, BG_t + \omega_{i,s}^{BV}\, BV_t + e_{it}$

For $t = s - 36, \ldots, s - 1$ and $s = 37, 49, 61 \ldots, 157$

Allocate the next 12 month returns from fund i to portfolio j that has the maximum one or two weights, $\omega_{i,s}^{j}$. This gives us portfolio returns, C_t^{j} for:

(i) Method C4: $j = SG, SV, BG, and\ BV$

(ii) Method C9: $j = SG, SV, BG, BV, Small, Big, Value, Growth, and\ Mixed$

Finally, performance evaluation is carried out as follows:

(i) Relative Style performance: $AR_t^{SG} = C4_t^{SG} - SG_t$

(ii) Regression Style performance: $C4_t^{SG} - R_t^{f} = \hat{\alpha} + \hat{\beta} SG_t + \hat{e}_t$

4.1.3. Dominant (Constant) Style

In this approach we assume that each fund has a single dominant style throughout the sample period. Accordingly, a unit trusts' investment style is identified as an average changing style over the period from 2005 to 2017. In contrast to the 'continuous changing style' approach, the allocation is done once on the basis of the average weight for each fund across the (time) sample. Specifically, 36-month rolling window is employed to obtain $\omega_{i,s}^{SG}, \omega_{i,s}^{SV}, \omega_{i,s}^{BG}$, and $\omega_{i,s}^{BV}$ in each year s for each fund i. Then the average value of the factor loadings is calculated over the whole sample period.

$$\overline{\omega}_i^j = \frac{1}{T}\sum_s \omega_{i,s}^j \quad for \quad j = SG, SV, BG, BV \tag{4.6}$$

Next, we construct four portfolios of stylized unit trusts based on the average highest factor exposure produced by the RBSA regression. We call this approach D4. Thereafter, average unit trusts relative performance is evaluated by comparing the stylized portfolio returns with their corresponding investment style mimicking portfolios. Unfortunately, there were little to no funds with factor exposure greater than 0.5, hence we were only able to perform the first method.

Summary: Dominant Style

For each fund $i = 1,2, \dots, N$, $t = s-36, \dots, s-1$ and $s = 37, 49, 61 \dots, 157$ find the weights $\omega_{i,s}^{j}$ from the regression

$$R_{it} = \omega_{i,s}^{SG}\, SG_t + \omega_{i,s}^{SV}\, SV_t + \omega_{i,s}^{BG}\, BG_t + \omega_{i,s}^{BV}\, BV_t + e_{it}$$

Calculate the average weight for each fund i across time.

Allocate the whole set of returns from fund i to portfolio j that has the maximum average weights, $\omega_{i,s}^{j}$. This gives us portfolio returns, $D4_t^j$ for: $j = SG, SV, BG, and\ BV$

Finally, performance evaluation is carried out as follows:

(i) Relative Style performance: $AR_t^{SG} = D4_t^{SG} - SG_t$

(ii) Style regression performance: $D4_t^{SG} - R_t^f = \hat{\alpha} + \hat{\hat{\beta}}(SG_t - R_t^f) + \hat{e}_t$

4.2. Event Time

Having settled on an appropriate benchmark to measure fund's expected return, there are several ways of proceeding to estimate short and long-run abnormal returns. In this section, we continue exploring return performance of UK equity funds by closely examining funds' investment style, using the alternative approach of event studies. Event studies are conducted to detect short-run and long-run abnormal performance in the UK equity funds. This usually requires the application of event time and calendar time approaches based on several benchmarks, namely, reference portfolio, control firm portfolio, and asset pricing model. In contrast to traditional event studies which employ specific events (such as IPO, earning announcement, etc.), the abnormal performance is examined based on the aggregate returns of funds' investment style. In particular, the performance is measured according to the hypothesized event of whether the style-adjusted performance of UK equity funds generate significant abnormal returns over a period of one to five years. Thus, abnormal returns are calculated based on an investment strategy that could be achieved by systematically buying units in funds with specific investment style objectives over an investment horizon of a one- to five-year period.

4.2.1. Event Time Portfolio Formation

The Return Based Style Analysis (RBSA) employed earlier in this chapter is used to identify funds' investment style. Every five years from 2005 to 2017, funds are regressed on an individual basis, before being grouped into four stylized and equally weighted portfolios based on their factor exposures over a three-year rolling window after fund inception dates. Specifically, the Small Growth (SG) portfolio represents the returns of equally weighted unit trusts with the highest exposure to small-growth stocks. The Small Value (SV) portfolio represents the returns of equally weighted unit trusts with the highest exposure to small-value stocks. The Big Growth (BG) portfolio represents the returns of equally weighted unit trusts with the highest exposure to big-growth stocks. The Big Value (BV) portfolio represents the returns of equally weighted unit trusts with the highest exposure to big-value stocks. The choice of an equally weighted over a value weighted portfolio is dictated by two reasons. First, data on fund size is not easily attainable in the UK mutual fund industry. Second, investors usually allocate their wealth equally across assets using the naive 1/N rule as explained by Benartzi and Thaler (2001).

Within each investment style category, the returns of each fund i is tracked over a five-year investment horizon. This is consistent with the monthly returns earned on a mutual fund that does not change its investment style for a five-year period. According to our sample, February 2005 represents the first hypothesized event date “investment style”. After this date, funds’ exposure to certain investment style are estimated over the prior three-year period (i.e., February 2002 to February 2005). We assume that investors were born in February 2005, and the returns of their preferred investment style is tracked for the next 12, 24, 36, 48, and 60 months. In February 2010, funds’ investment style is re-estimated, and similar steps are applied to track their style performance for the next five years. Thus, the official start date is February 2005 and we reset the clock every five years. For example, Figure 4.1 indicates that, between February 2002 and February 2005, fund 1(Rp1) has the highest factor exposure to small-growth mimicking portfolio among the other style mimicking portfolios. Therefore, Rp1 subsequent five years’ return is allocated to small-growth stylized portfolio. Similarly, small-growth mimicking portfolio has the highest factor exposure between February 2008 and February 2010 rolling window, and Rp1 subsequent 5 years’ return is allocated to small-growth stylized portfolio. Hence, Rp1 has a consistent small-growth investment strategy throughout the studied period.

For a fund that is incepted during the test period, the fund was only included in the data set if it had at least 48 months of returns information for that period. Thus, factor exposure across the four style mimicking portfolios can be estimated over a three-year period then tested for the subsequent 12 months. For example, if a fund’s inception date is December 2005, then in December 2008, it will be assigned to one of the four stylized investment portfolios and its return will be tracked over the next five years to December 2013. This is the example of SG(Rp3) in Figure 4.1.

Another problem we encountered is that, if a fund ceases to exist (because of, say, closure or merger) during the investment horizon, then we replace the returns of that fund by the benchmark return. This is the example of SG(Rp49) in Figure 4.1, where the red arrow represents the replacement of fund return by the benchmark return. Mitchell and Stafford (2000) observed that using the benchmark return might create an upward bias in the estimated BHAR returns if a fund's delisting is caused by bankruptcy. However, most defunct funds in our sample have marginally preserved their value and were not insolvent. By preserving value, we mean the sale of all fund assets and distribution of the proceeds to shareholders. There is only one conventional fund that was bankrupt, with its Return Index (RI) dropping from 187 to 0.18. However, this fund has returns information of less than 48 months and was excluded from the data set. Finally, long-term unit trusts' abnormal performance is evaluated using event time approaches. Thus, the following null hypotheses are tested: the cross-sectional means of (i) buy-and-hold abnormal return, and (ii) cumulative abnormal return is zero over investment horizons of a one to five-year period.

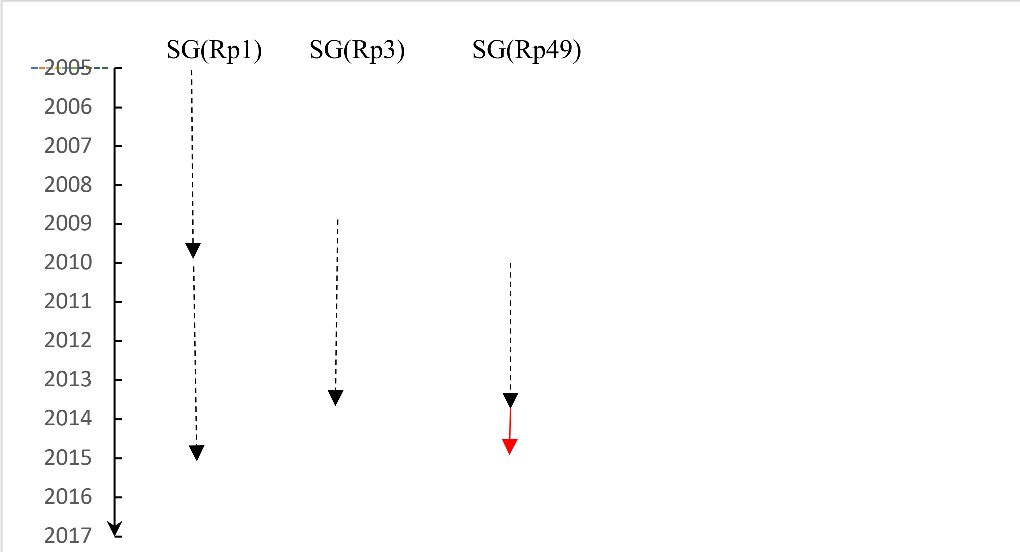


Figure 4.1: An example of Small Growth (SG) investment style.

4.2.2. Event Time Measures

4.2.2.1. Buy and Hold Abnormal Returns (BHAR)

Following the standard methodology outlined in Lyon, Barber, and Tsai (1999), BHARs is calculated as

$$BHAR_{i,\tau} = \left[\prod_{t=1}^{\tau}(1 + R_{i,t})\right] - \left[\prod_{t=1}^{\tau}(1 + R_{it}^{b})\right] \tag{4.7}$$

where τ is the period of investment in months following the event month, and $R_{i,t}$ is the return on unit trust i within a given investment style in month t. The benchmark return, R_{it}^{b} is formed using a reference portfolio that captures the funds' expected return. Reference portfolios are constructed in two ways. The first is the factor mimicking portfolios of size, and value-growth orientation, as being described earlier; Small-growth (SG), Small-Value (SV), Big-Growth (BG), and

Big-Value (BV) mimicking portfolios were obtained from Exeter University, Centre for Finance and Investment (Gregory et al., 2013).[9] Here, we call these indices characteristic-based reference portfolios. Although, the return of characteristic-based reference portfolios might not perfectly reflect the expected return of unit trusts (Fama, 1998), bad-model problems are also extremely problematic. These characteristic-based portfolios are therefore expected to capture the cross-sectional and time-series variation in mutual funds as closely as possible. Besides, the characteristics-based reference portfolios are constructed using the FTSE 350 firms. Hence, they represent an appropriate investable opportunity set as advocated by Liu and Strong (2008). It is worth noting that our characteristics-based reference portfolio gets closer to the control firms approach for measuring abnormal performance in event studies. The second reference portfolio measurement is the value-weighted market portfolio (FTSE100). Despite the dominant use of multifactor models in academic studies, it is common among practitioners to evaluate funds' return against a general market index (Betker and Sheehan 2013).

The BHAR derived from these reference portfolios represents the abnormal return on an equally weighted stylized portfolio compared to that of an equivalently style-controlled passive investment portfolio with yearly rebalancing. The result assumes that unit trust investment style is consistent throughout the whole test period (5 years after the event date). Barber et al. (1999) documented that, long-run abnormal return calculated using the BHAR approach yields mis-specified test statistics,

[9] The reader is referred to Gregory et al. (2013) for more details on portfolios construction.

which in turn would cause rejection of the null hypothesis when it is true. The power of the standard test statistic suffers from positive bias introduced by new listing or survivor bias, while rebalancing and skewness creates a negative bias in test statistics. However, these misspecification problems are at least partially mitigated in our study. The characteristics-based reference portfolios deployed to capture expected returns are rebalanced on a yearly basis[10], whilst the FTSE 100 is rebalanced quarterly. This is likely to match managers' rebalancing strategy. Furthermore, when a mutual fund ceases to exist (i.e., dead, suspended, merger, or delisted) within the 5-year measurement period, the dead fund's return is replaced by a benchmark return. This would mitigate the survivorship bias as advocated by Mitchell and Stafford (2000). Additionally, both the characteristics-based reference portfolio and unit trusts returns are well diversified, hence one would expect their returns to be normally distributed. The Central Limit Theorem implies that the "the sum of a large number of independent random variables has a distribution that is approximately normal" (Ross, 1976). Nevertheless, Brav (2000) and Mitchell and Stafford (2000) pointed out that the long-horizon return observations in event portfolios are cross-correlated, since co-movements amongst securities returns are widely observable. Thus, the lack of independence across the unit trusts' return will lead to a severe misspecification of the test.

[10] See Gregory et al. (2013).

Summary: Event Time Approach

For each fund $i = 1,2, ..., N$, find the weights $\omega_{i,s}^{j}$ from the regression

$R_{it} = \omega_{i,s}^{SG}\ SG_t + \omega_{i,s}^{SV}\ SV_t + \omega_{i,s}^{BG}\ BG_t + \omega_{i,s}^{BV}\ BV_t + e_{it}$

For $t = s - 36, \dots, s - 1$ and $s = 37,38\dots127$

Allocate the next 12,24,36,48, and 60-month returns from fund i to portfolio j that has the maximum weights, $\omega_{i,s}^{j}$.

Repeat the process every 5 years to construct portfolio j returns as $(1+R_{it}^{j})(1+R_{it}^{j})\dots\dots(1+R_{nt}^{j})$ for $j = SG, SV, BG, and\ BV$

Finally, performance evaluation is carried out as follow:

Aggregate Abnormal Returns: $BHAR_{i,\tau} = [\prod_{t=1}^{\tau}(1 + R_{i,t})] - [\prod_{t=1}^{\tau}(1 + R_{it}^{b})]$

4.2.2.2. Cumulative Abnormal Average Returns (CAARs)

Although the BHAR methodology seems to be more representative of investor experience, this section briefly reports the cumulative abnormal return (CAR) results for a completeness and robustness check. The CAR approach is used to inspect the null hypothesis of zero mean cumulative abnormal returns and is calculated as:

$$CAR_{i,\tau} = \sum_{t=0}^{\tau-1} AR_{i,t} \tag{4.8}$$

$$AR_{i,t} = R_{i,t} - E(R)_{i,t}) \tag{4.9}$$

where $AR_{i,t}$ is the abnormal return of an equally weighted portfolio of unit trusts with investment style i, and τ is the investment horizon. In this scenario, unit trusts are grouped into four stylized equally weighted portfolios based on their highest factor exposure to RBSA regression. $R_{i,t}$ is the realised unit trust i return within a given investment style at time t, and $E(R_{i,t})$ is the expected return on style i at time t, which is calculated using either the characteristic-based reference portfolios or the market index (FTSE 100). However, drawing inferences on long-term performance from such a procedure remains challenging. Thus, abnormal returns over long-term horizons are highly sensitive to benchmark selection, and abnormal returns will always be measured with error due to the lack of a perfect model of expected returns.

4.2.2.3. Test statistics in event time

Barber et al. (1999) controlled for cross-correlation and skewness biases using bootstrapped skewness-adjusted t-statistics. This verdict largely motivates the use of Johnson's (1978) skewness correction approach. Whereby, the BHAR's conventional t-statistic is calculated as:

$$t_\tau = \frac{\overline{BHAR_\tau}}{\sigma(BHAR_\tau)/\sqrt{N}} \tag{4.10}$$

where $\overline{\mathrm{BHAR}_\tau}$ is the cross-sectional sample mean, and $\sigma(\mathrm{BHAR}_\tau)$ is the cross-sectional standard deviation, and $\sqrt{N}$ is the number of unit trusts within the investment style during investment horizon τ.

Similarly, the CAR's test for significance is given by:

$$t_{\tau} = \frac{\overline{CAR_{\tau}}}{\sigma(CAR_{\tau})/\sqrt{N}} \tag{4.11}$$

where $\overline{CAR_{\tau}}$ is the cross-sectional average of individual fund CAR within a given style, and $\sigma(CAR_{\tau})$ is the standard deviation for the same group of funds CARs, and $\sqrt{N}$ is the number of unit trusts within the investment style during investment horizon τ. Apart from the conventional student's t-test, we also test for robustness using a nonparametric test, which will be explained in section (4.4). We apply baseline and wild-adjusted bootstrap procedures to deal with non-normally distributed returns data in event studies.

4.3. Calendar Time Methodology

4.3.1. Calendar Time Abnormal Return (CTAR)

In this section, we continue exploring the return performance of UK equity funds using the calendar time approach. Specifically, we examine the propensity of the style-adjusted performance of UK equity funds to generate statistically and economically significant abnormal returns over a period of five-year. The calendar-time portfolios are constructed from sample funds that have experienced the hypothesized events in the prior five-year period. The hypothesized event is funds' investment style, which is identified based on the highest factor exposure produced by RBSA regression. Consequently, the returns of the calendar-time portfolios can be achieved by systematically buying units in funds with a specific investment style objective over an investment horizon of a five-year period. The calendar-time approach sums up the event funds' returns on the biases of equally weighted returns in a calendar month. In any given month, each fund that has experienced the event in the previous 60 months is included in the portfolio. The weight given to that fund is $1/N_t$, where N_t is the number of funds in that portfolio in month t. Given the fact that our data sample started in February 2002, February 2005 represents the first hypothesized event, and funds that experience the event (i.e., those born in the previous 3 years and that have at least 12 months of observations) are included within one of the four investment style categories (calendar time portfolio). Thus, we assume that investors were born in February 2005, and the returns of their preferred investment style are traced for the subsequent 60 months. In February 2010, funds' investment styles are re-estimated, and similar steps are applied to track their style performance for the next 5 years. Thus, the official start date is February 2005, and we reset the clock every 5 years.

Figure 4.2 gives an example of the formation of small-growth calendar time portfolio, where four events took place between 2005 and 2017. In detail, in February 2005, fund A is identified as small growth (i.e., the highest loading in RBSA regression is $\omega_{i,s}^{SG}$), and the subsequent 60 months of fund A's returns are allocated to small-growth calendar time portfolio.

Then, in January 2007, fund B experienced the small growth event and small-growth calendar time portfolio contains two funds at this point (A and B). In February 2010, the portfolio drops fund A since it joined the portfolio 60 months ago. There are two possible scenarios for fund A; either fund A's style exposure has shifted over the testing period, or fund A contains less than 12 months of return observations beyond February 2010. In July 2011, fund C entered the small-growth calendar time portfolio. While fund B completes its 60 months in January 2012, it remains exposed to a small growth investment style[11], and therefore immediately re-enters the small-growth calendar time portfolio in the same month.

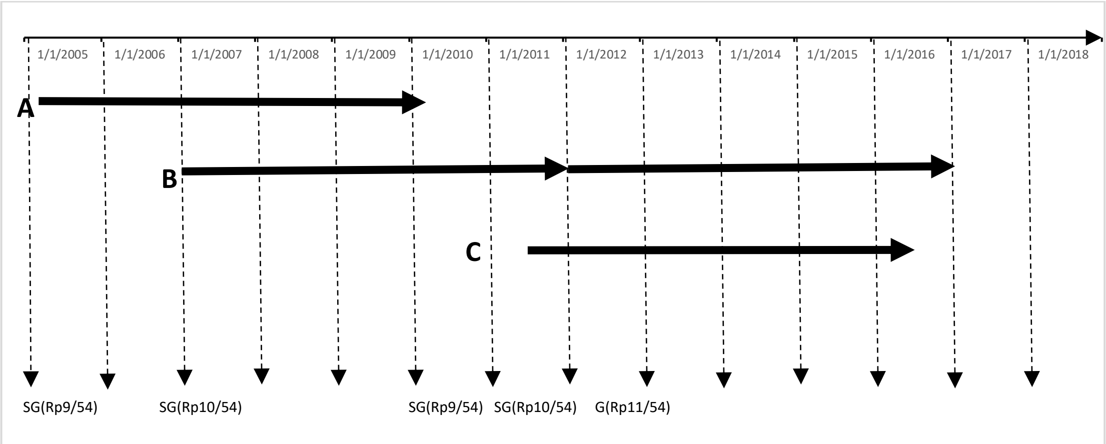


Figure 4.2: An example of monthly small-growth calendar time portfolio.

Therefore, the monthly return of the small-growth calendar time portfolio is computed as follows:

$$R_{\tau,t}^{SG} = \sum_{i=1}^{n_{\tau,t}} w_{i,t} R_{i,t} \tag{4.12}$$

where $w_{i,t}$ is $\frac{1}{n_{\tau,t}}$, and $n_{\tau,t}$ is the number of funds in small-growth calendar time portfolio in month t for $\tau = 60$. In effect, the small-growth calendar time portfolio represents a strategy of investing a fixed amount of cash in small-growth funds for 60 months before selling it and reinvesting the proceeds in small growth funds for

[11] For example, between January 2009 and January 2012, the highest loading in RBSA regression for fund B is $\omega_{i,s}^{SG}$.

another 60 months. Due to the small number of funds in our data sample, particularly of ethical funds, the construction of calendar time portfolios based on shorter investment horizon (i.e., $\tau = 12, 24, 36,$ and 48) is not possible. Thus, our results report the calendar time abnormal returns based on investment strategy that could be achieved by systematically buying units in funds with a specific investment style objective for a period of 60 months from the event (investment style).

As opposed to the raw returns used in the calendar time portfolio, the returns of interest are the abnormal returns in the long term. The main objective is to test the null hypothesis of no abnormal returns for a holding period of 60 months. Thus, if the null hypothesis is true, then the calendar time abnormal return for any month is expected to be zero. The standard way of calculating the monthly calendar time abnormal returns is the following:

$$R_{\tau,t}^{j} - r_f = \alpha + (R_{\tau,t})^E + \varepsilon_t \qquad (4.13)$$

$R_{\tau,t}^{j}$ is the monthly excess return on portfolio of event funds, where $j = SG, SV, BG, BV$. r_f is the 1-month Treasury bill (risk-free) rate of return. α is Jensen's alpha that measures the calendar-time portfolio's abnormal monthly mean return over investment horizon of $\tau = 60$. $(R_{\tau,t})^E$ is the expected return on the event portfolio, and ε_t is the error term which is assumed to follow a normal distribution with zero mean and constant variance. The debate of which model is best suited to capture long term expected returns remains unresolved. For example, Fama (1998, p.291) stated that "all models for expected returns are incomplete descriptions of the systematic patterns in average returns." This is because long-term abnormal return is sensitive to the choices of models employed for evaluating the theoretical expected returns, while it also exhibits non-normal residual distributions.

We therefore formed the $(R_{\tau,t})^E$ in two ways. First, using a reference portfolio that captures the size, and value-growth orientation of calendar-time portfolio. The characteristic-based reference portfolios were obtained from Exeter University's Centre for Finance and Investment (Gregory et al., 2013). Four portfolios are considered. SH represents the returns on value weighted portfolio of small stocks with high book to market ratio. SL represents the returns on value weighted portfolio of small stocks with low book to market ratio. BH represents the returns on value weighted portfolio of big stocks with high book to market ratio. Finally, BL represents the returns on value weighted portfolio of big stocks with low book to market ratio to reflect the return on style benchmark indices:

$$(R_{\tau,t})^E = R_{ft} + \beta(R_{bt} - R_{ft}) \tag{4.14}$$

Accordingly, we assume that the return of a characteristics-based reference portfolio perfectly reflects the expected return of calendar-time event funds' portfolio. This approach was advocated by Angelidis et al. (2013), who argue that the use of market benchmarks rather than fund's self-designated benchmark biases the performance evaluation processes. However, our characteristic-based reference portfolio ignores risk factors beyond the size and value-growth characteristics and those may not be completely able to describe the cross-section variation of expected returns.

In the second formation, we employ a regression-based model where the market model capital asset pricing model (CAPM), and the Fama and French 3-factor model are used to measure the expected returns.

$$(R_{\tau,t})^E = R_{ft} + \beta(R_{mt} - R_{ft}) \tag{4.15}$$

and

$$(R_{\tau,t})^E = R_{ft} + \beta_{Rm}[\mathrm{R_{mt}} - \mathrm{R_{ft}}] + \beta_{\mathrm{SMB}}(SMB)_t + \beta_{\mathrm{HML}}(HML_t) + \varepsilon_t \tag{4.16}$$

where R_m is a proxy for the return on the market portfolio, R_f is the 1-month Treasury bill (risk-free) rate of return, SMB is the difference in returns between portfolios made up of small and big stocks in calendar month t , and HML is the difference in returns between portfolios made up of stocks with high and low book-to-market ratios in the calendar month t.

Despite the popularity of the CAPM among practitioners, the model ignores risk factors beyond the market portfolio, and the multifactor factors model is still dominant in academic studies (Betker and Sheehan, 2013). On the other hand, Lyon et al. (1999) favour the use of simple characteristic-based reference portfolio and argue that the common regression-based models have systematic problems in explaining the expected returns. For example, the three-factor model wrongly assumes that there is a linear relationship in factor exposures. Furthermore, although the module predicts a strong return for growth stocks over value stocks, the return is most noticeable for small size stocks (Loughran, 1997). Thus, the three-factor model ignores the intersection between factors. Additionally, Al-Horani et al. (2003) argue against the three-factor model since it may not be completely able to describe the cross-sectional variation of expected returns. For example, within the UK context, they showed that including a R&D expenditure factor in estimating risk premia can significantly improve the explanatory power of the three-factor model. More recently, Pettengill et al. (2014) argue that the three-factor model is biased against portfolios of value-oriented stocks. This is because value-oriented portfolios have positive HML loading, which will inherently increase their expected returns whilst reduce their estimated abnormal returns α's.

4.3.2. Test statistic in calendar time

Assuming that the model of expected returns is correct, the time series of equally weighted event funds' portfolios are estimated using equation (4.13). The intercept (α) derived from this regression indicates whether or not the mean abnormal return is statistically different from zero. The ordinary least squares (OLS) provides the best linear unbiased estimators (BLUE) when the Gauss Markov's assumptions are satisfied, such as homoscedasticity, non-stochastic regressor, and uncorrelated disturbance terms. Nonetheless, the regression suffers from heteroscedasticity since the number of event funds varies each month (Loughran and Ritter, 2000). Thus, the regression estimators are still unbiased, but they are no longer efficient. Several studies address the question of how to correct this bias. For example, Mitchell and Stafford (2000) suggested the use of weighted least squares regression, whereby the square root of the number of event funds in calendar time portfolio is used as weights. By taking an example of small-growth reference portfolio model of equation (4.13).

$$R_{\tau,t}^{SG} = \alpha + \beta\, SG_{\tau,t} + \varepsilon_t \tag{4.17}$$

where $R_{\tau,t}^{SG}$ is the monthly returns on portfolio of small-growth funds in τ holding period. SG is the characteristic-based reference portfolio and represents the return on value weighted portfolio of small stocks with high book to market ratio. Mitchell and Stafford (2000) assume the heteroscedasticity takes the form of $\sigma^2 = \omega^2$, where $\omega^2 = {}^{1}/_{n_{\tau,t}}$ and $n_{\tau,t}$ refer to the number of event funds in the calendar month t. Therefore, we can standardize the residual variance with the transformed regression.

$$\sqrt{n_t} R_{\tau,t}^{SG} = \sqrt{n_t}(\alpha + \beta\, SG_{\tau,t} + \varepsilon_t) \tag{4.18}$$

Hence, $Var\ (\varepsilon_t)/\omega_t = 1$. However, this transformation assumes that funds' residuals are independent, and the calendar time methodology loses its main intended objective. Alternatively, Mitchell and Stafford (2000) suggest the use of a non-parametric bootstrap procedure to capture the critical value. However, OLS inefficient estimators may persist, whereby the distribution of small-growth reference portfolio does not replicate the covariance structure of the calendar time event funds' portfolio. The risk of characteristics-based reference portfolio does not perfectly reflect the risk of calendar-time event funds' portfolio. Therefore, the residuals may suffer from autocorrelation when the sample is tilted toward an unobservable common factor.

Another alternative is to use Feasible Generalized Least Squares (GLS) as proposed by Gregory, Guermat and Al-Shawawreh (2010). In their paper, the variance takes the form of a linear function of the number of funds in the calendar-time portfolio in each month. The procedure is carried out as follows: first they obtained the unrestricted disturbance using equation (4.17). Thereafter, they estimated the regression

$$\log(\hat{\varepsilon}_t^2) = \delta_0 + \delta_1 \log(n_t) + u_t \tag{4.19}$$

where n_t refer to the number of event funds in the calendar month t. Then, they set the projected variance $\widehat{Var_t}(\varepsilon_t) = exp(\hat{\delta}_0 + \widehat{\delta}_1 n_t)$. The authors compare their result with the covariance matrix estimator that is proposed by White (1980). While White's estimator is consistent under heteroscedasticity of unknown form, it can nevertheless be quite biased when the sample size is small. Gregory et al. (2010) concluded that the Feasible GLS delivers similar standard errors as in the OLS with robust White's variance estimators, but has a better adjusted-R-square. Accordingly, we believe OLS with robust White's and Gregory et al.'s Feasible GLS might be the most appropriate tests to reduce misspecification in tests of long-term calendar time method.

4.4. The Bootstrap Procedure: Skill versus Luck in Performance.

Assessing investment fund performance is an important theoretical and empirical issue in finance. The magnitude of invested funds and the fees imposed by investment managers have drawn a large number of studies on the performance of investment funds. However, testing fund performance remains problematic despite decades of research. Among the challenges confronting researchers are that of measuring return, as discussed in earlier chapters, and the bad model problem (Fama, 1998). This section focusses on a further challenge, namely, data snooping. This line of research argues that tests of fund performance often ignore the fact that some funds over-perform (under-perform) because of (bad) luck rather than (poor) skill. Separating out luck and skill is therefore the aim of this section.

Most of the existing studies on modelling the role of luck in funds' performance have focused on testing the persistence of funds' returns. Thus, managerial skills were directly related to whether past winners (losers) continue to produce superior (inferior) performance (Fletcher, 1997; Brown, Draper and McKenzie, 1997; Quigley and Sinquefield, 2000; Fletcher and Forbes, 2002). However, persistence tests have many weaknesses. For example, Sullivan et al. (1999) raised the concern of potential data snooping issues. They defined data snooping as a situation whereby the same set of data is used to construct trading rules and to test them. Therefore, if enough trading rules are tested, some of them would generate significant results, solely due to chance. Fama and French (2010) noted that ranking funds according to their short-term past performance is subject to noise. Hence, little significant evidence of persistence can be drawn under such a methodological framework. Aside from the possibility that luck can also persist, Kosowski et al. (2006) showed that the standard statistical significance test of abnormal fund performance may give misleading inferences. In particular, the standard test statistic imposes unrealistic assumption about the ex-ante distribution of funds' returns. For example, the parametric tests require the abnormal returns measure to be normally distributed. Although, the central limit theorem implies that equally weighted portfolio approach normality regardless of the individual stocks' distribution, mutual funds tend to hold few stocks or allocate their holding in few industries. Similarly, returns on factor loadings may not be normally distributed and the co-skewness between these factors and fund's returns may persist. Active managers tend to change their risk appetite in accordance with market condition and their relative performance. Thus, even if individual fund returns are normally distributed, the cross-sectional distribution of the alphas may be non-normal. These conditions can contribute to the non-normality of abnormal mutual fund returns, hence inducing miss-specified inferences of the standard test statistics and jeopardizing the power of asset pricing models to detect abnormal performance. Consequently, we apply bootstrap procedures to determine whether the significant performance estimates are the result of superior/inferior fund managers' skill or simply due to extraordinarily good/bad luck.

4.4.1. Baseline Bootstrap

Following the bootstrap method of Kosowski et al. (2006), we analyse the distribution of individual fund's performance as generated by commonly used performance models, such as the single factor model, and the three-factor model. We assume that returns are generated by the following general process

$$R_{i,t} = \alpha_i + \beta_i X_t + \varepsilon_{i,t} \tag{4.20}$$

where $R_{i,t}$ is the excess monthly return on fund i at time t, α is Jensen's alpha, which measures monthly abnormal return, X_t is a matrix of risk factors, β_i is a vector of loadings, and $\varepsilon_{i,t}$ is the error term. Using the ordinary least squares (OLS) method, Jensen's alpha, the factor loadings, and the residuals are estimated and saved using monthly excess returns for each fund.

Then, for each fund, we randomly re-sample with replacement from the saved residuals $\hat{\varepsilon}_{i,t}$ for $i = 1,2,..n$ funds of length T_i. The (time) sample starts with the inception date and continues until the end of the data collection period, so $T_{i,1}$ is specific to fund i in calendar time. Next, we construct a sample of monthly pseudo excess returns $(R_{i,t})^b$ using the resampled bootstrap residuals, alongside the original chronological ordering of the explanatory factors X_t, while imposing the null hypothesis of zero abnormal returns.

$$(R_{i,t})^b = \hat{\beta}_i X_t + \hat{\varepsilon}_{i,t}^b \tag{4.21}$$

Thus, for each fund i, pseudo time series returns $(R_{i,t})^b$ are created with zero alpha by construction, where $b = 1\ to\ 1000$ and represent the b^th bootstrap. Next, the simulated returns $(R_{i,t})^b$ are regressed on factor models to generate alpha luck distribution and its corresponding t-statistic.

$$(R_{i,t})^b = \alpha_i + \beta_i X_t + \varepsilon_{i,t} \tag{4.22}$$

Thus, for each of the n funds in our sample, we estimate and save the respective alpha and its corresponding t-stat in matrices of dimension $n \times b$ as follows.

$$\begin{bmatrix} a_1^1 & a_1^2 \dots \dots & a_1^{1000} \\ a_2^1 & a_2^2 \dots \dots & a_2^{1000} \\ a_n^1 & a_n^2 \dots \dots & a_n^{1000} \end{bmatrix}$$

$$\begin{bmatrix} t_1^1 & t_1^2 \dots \dots & t_1^{1000} \\ t_2^1 & t_2^2 \dots \dots & t_2^{1000} \\ t_n^1 & t_n^2 \dots \dots & t_n^{1000} \end{bmatrix}$$

Next, the estimated a_i^b and t_i^b for $i = 1, \dots., n$ and $b = 1, \dots, 1000$ are ranked from the highest to the lowest to form the luck distribution (cumulative distribution function of the alphas and its correspondent t-stats) under the null hypothesis of zero abnormal returns. Thus, the first row of each matrix represents the highest possible sampling variation around a true value of zero alpha and are entirely due to extraordinary good luck. While the bottom row of each matrix contains the lowest possible sampling variation around a true value of zero alpha and are entirely due to extraordinary bad luck. Finally, the alphas and their t-statistics that were estimated in the first step are ranked and then compared with their luck distributions. For example, if the highest percentile t-statistic of alpha (top 1%) of the original distribution exceed the highest percentile t-statistic of alpha from the bootstrap distribution (luck distribution), then we reject the null hypothesis that its performance is due to luck. Hence, we conclude that sampling variation is not the source of a high t-statistic of alpha but rather that genuine stock-picking skills exist. This can be repeated for any quantile in the performance distribution, for example on the left tail of the distribution we interpret whether fund managers possess bad skills, or whether the underperformance is solely due to chance (bad luck).

However, the baseline bootstrap proposed by Kosowski et al. (2006) assumes independence of the residuals, while the risk factors remain constant across the sample period. Hence, the baseline bootstrap ignores systematic risk and assesses fund manager's skills in terms of non-systematic risk. In other words, it ignores the correlation of the abnormal returns across funds when the equilibrium model does not capture all common variation in fund returns (Blake et al., 2014).

Another limitation is that during the random sampling from individual fund residuals, fund returns lose any properties of serial correlation in the simulation runs. However, as robustness checks, Kosowski et al. (2006) performed simulations in block lengths corresponding to the suspected order of serial correlation, where a

Lagrange Multiplier test is used to select the order of autocorrelation. They also corrected all t-statistics based on Newey-West autocorrelation and heteroscedasticity adjusted standard errors. They concluded that the results change very little, whether the bootstrap procedure is adjusted to incorporate block lengths of the suspected order of serial correlation or Newey-West autocorrelation and heteroscedasticity adjusted standard errors.

Fama and French (2010) argued that the baseline bootstrap is biased towards positive performance. Consequently, they proposed a bootstrap procedure that preserve the common dependency between explanatory factors and residuals. Specifically, abnormal fund returns were estimated from the observed fund returns. Then monthly pseudo zero alpha returns were constructed by jointly re-sampling fund and explanatory returns. Furthermore, fund's inclusion criteria were restricted to funds that have a minimum of 8-month observations rather than the 60-month time period used in Kosowski's baseline bootstrap. Thus, Fama and French's bootstrap is less subject to survival bias.

Nonetheless, we favour Kosowski over Fama and French's bootstrap approach since funds do not all exist at the same time. Indeed, the baseline residual bootstrap ensures that the number of return observations in the data generation process always matches the fund's actual number of return observations. Furthermore, our analysis is restricted to funds that have at least 36-month of observations. This is because funds with fewer observations are more likely to have a higher variability in their alpha estimates, thus widening the extreme tail of luck distribution and leading to inferences that are biased toward luck performance. We also favour the t-statistic of alpha, t_{α}, rather than α as the measure of fund performance since it has better statistical properties and controls for the variability in funds return. Mamaysky et al. (2007) showed that when sorting funds according to their alpha, the top and bottom deciles contain funds having the greatest estimation error rather than the best/worst performing funds. Thus, our study adopts the t-statistic of alpha as the fund performance measure.

Finally, in order to eliminate the effects of the correlation variability in the explanatory factors returns and residuals, we considered the single and the three-factor models; since such cross-sectional dependency is mostly driven by a misspecification in the performance model. Furthermore, our baseline bootstrap procedure is applied across funds of different investment styles, namely, small-growth, small-value, big-growth, and big-value investment styles. By doing so, we account for homogeneous risk across funds, which might not be captured by the asset pricing model. We also propose an alternative bootstrap procedure, namely, the wild-adjusted bootstrap that mimics the true funds' performance distribution. Thus, our results are robust with respect to the variability of the bootstrap performance estimates.

4.4.2. Wild Adjusted t-statistic Bootstrap

The baseline bootstrap is designed to distinguish skill from luck when fund returns are independently and identically distributed. However, Wu (1986) and Beran (1986) demonstrated that the scheme does not work well in the presence of heteroscedasticity. The data generation process of the standard bootstrap procedure cannot mimic the heteroscedasticity inherent within the parent distribution. White (1980) proposed one of the earliest works to correct for inference in the presence of unknown form of heteroskedasticity. Nonetheless, MacKinnon and White (1985) showed that the heteroskedasticity consistent covariance matrix estimator can be seriously biased, when the distribution is drawn from a small-sample size or exhibits high residuals. Accordingly, and in light of the work of Gregory et al. (2010), we perform a wild-adjusted bootstrap procedure that mimics and incorporates the characteristics of the true fund's returns distribution. Our wild adjusted bootstrap has similar steps to those described in the baseline bootstrap. However, the refinement to retain the characteristics of the parent distribution involves three main steps:

First, for each of the 1000 bootstrap replications, the bootstrap residuals in equation (4.21), $\hat{\varepsilon}_{i,t}^{b}$, are obtained as the product of the original residuals, $\hat{\varepsilon}_{i,t}$ (equation 4.20), and an independent random variable $\eta_{i,t}$. Following the proposition of Davidson et al. (2007), $\eta_{i,t}$ is defined as:

$$\eta_{i,t} = \begin{cases} 1 & w.p. \quad p = \dfrac{1}{2} \\ -1 & w.p. \quad 1-p \end{cases} \tag{4.23}$$

Therefore, the resampling residuals distribution has identical distribution to the original residuals in term of mean, and variance.

Whereby, $E(\hat{\varepsilon}_{i,t}^{b}) = E(\eta_{i,t})E(\hat{\varepsilon}_{i,t}) = 0$, and $V(\hat{\varepsilon}_{i,t}^{b}) = V(\eta_{i,t})V(\hat{\varepsilon}_{i,t}) = V(\hat{\varepsilon}_{i,t})$

Second, the t-statistic of the alphas of the parent distribution (equation 4.20) are adjusted using Johnson's (1978) skewness adjusted t-statistics approach. The standard t-statistic is calculated as

$$tstat_i = \frac{\hat{\alpha}_i}{SE(\hat{\alpha}_i)} \tag{4.24}$$

where $\hat{\alpha}_i$ is the estimated alpha, and $SE(\hat{\alpha}_i)$ is the standard error of alpha for each fund i . The above standard t-statistic is adjusted for skewness using the following formula:

$$SKadj - tstat = \sqrt{N}\left(S + \frac{1}{3}\,\hat{\gamma}\,S^2 + \frac{1}{6N}\hat{\gamma}\right) \tag{4.25}$$

where $S = tstat/\sqrt{N}$, and $tstat$ is the standard t-statistic given in (4.24). $\hat{\gamma}$ is the coefficient of skewness of the regression residuals. We are therefore able to construct a nonparametric performance distribution across funds that mimics the actual distribution of funds' performance, assuming that Johnson's (1978) skewness correction is adequate. Finally, the empirical distribution of the bootstrapped alphas, standard t-statistics, and adjusted t-statistics are used to assess the significance of their initial counterparts given a pre-determined level of significance.

However, inferences drawn from such a procedure must be interpreted with caution for two reasons. First, similar to the baseline bootstrap, our bootstrap assumes serial independence in residuals. Thus, in the presence of autocorrelation in residuals, the simulation run will fail to imitate the parent distribution. However, Kosowski et al. (2006) adjusted the bootstrap t-statistics of alpha based on Newey-West autocorrelation adjusted standard errors. Furthermore, they performed a block bootstrap procedure (up to 10 months block) but their result changed very little. Second, cross-sectional correlation in funds' residuals represent a major challenge in our analysis. For example, suppose $COV(\varepsilon_1, \varepsilon_2) \neq 0$, where ε_1 and ε_2 are the residuals of fund 1 and fund 2. Then, under the independence assumption of our bootstrap, any effects of cross-fund correlations in returns will be lost in the data generation process. Although Fama and French bootstrap maintains the time ordering of residuals across all funds in each bootstrap, this requires funds to survive the whole sample period. As far as we know, cross-sectional correlation cannot be tackled in such a situation.

4.5. Conclusion

There is a continuous debate over which performance measures and time intervals are most suitable for fund's investors. Each methodology has its own pros and cons, showing different explanatory powers and highlighting a specific aspect of fund performance. This chapter comprehensively describes the methodologies used to evaluate UK-equity fund's performance based on funds' investment style. The UK-equity funds are split into 4 style investment categories and post returns are then used to evaluate future performance. This concludes the discussion of various performance measures based on style sorting rules along with many refinements. These performance measures suffer from fewer statistical flaws and better reflect investors experience from adopting such an investment strategy. Furthermore, it offers better information about funds managers' skills and their future performance.

Chapter 5

Style Analysis

5.0. Introduction

This chapter aims to address the following empirical questions. (i) What investment style is adopted across UK-equity funds? (ii) Is style investing profitable? (iii) Do ethical funds invest differently than their conventional counterparts and (iv) do they pay a higher price for their ethical consideration? We examine fund manager's stock selection behaviour and fund's investment policies and objectives in the context of returns-based style analysis. In particular, market capitalization and value-growth orientation are investigated to identify implicit links between investment styles and fund's performance. The main benefit of style analysis is to enable fund performance measurement by creating a generic benchmark that quantifies the performance more effectively than the market index. To this end, this chapter will discuss the empirical results of Sharpe's asset-classes model in the context of UK equity funds distinguishing between conventional and ethical funds.

5.1. Continuous changing style

5.1.1. Result based on continuous changing style (C4)

Table 5.1 presents the number of unit trusts in each stylized portfolio over a 36-month rolling window. It can be seen that the total numbers of equity funds have increased steadily over the studied period. The number of funds with small-cap exposure have upward trends up to 2008 then declined dramatically between 2008 and 2012. Similar behaviour can be observed in big-growth funds, but the decline continues up until 2014. Funds tilted towards big-value stocks have increased considerably between 2007 and 2013. In short, the results show that more than two third of the funds have a tendency to favour big oriented shares. It is clear that the bulk of unit trusts have shifted their preferences toward big–value stocks between 2008 and 2014. Meanwhile, investments in small cap companies dropped remarkably around the period of the 2008 financial crisis.

Table 5.1: Number of unit trusts in four stylized portfolios based on the distribution of estimated factor loadings for unit trusts using 36-month rolling window.

Year	*ROL*	*TNF*	*NSG*	*NSV*	*NBG*	*NBV*
2005	37	121	29 (24%)	20(17%)	69(57%)	3(2%)
2006	49	138	41(30%)	23(17%)	58(42%)	16(12%)
2007	61	154	47(31%)	37(24%)	68(44%)	2(1%)
2008	73	161	49(30%)	24(15%)	76(47%)	12(7%)
2009	85	167	12(7%)	28(17%)	104(62%)	23(14%)
2010	97	172	9(5%)	15(9%)	94(55%)	54(31%)
2011	109	175	11(6%)	14(8%)	64(37%)	86(49%)
2012	121	178	11(6%)	6(3%)	35(20%)	126(71%)
2013	133	179	22(12%)	21(12%)	20(11%)	116(65%)
2014	145	180	26(14%)	27(15%)	20(11%)	107(59%)
2015	157	184	54(29%)	12(7%)	45(24%)	73(40%)
2016	169	181	41(23%)	7(4%)	130(72%)	3(2%)

Notes: The symbols used to denote the investment style imply the following: ROL = month, TNF= total number of studied funds, NSG = number of small-growth, NSV= number of small-value, NBG= number of big-growth, and NBV= number of big-value investment style.

Table 5.2 shows the descriptive statistics of four equally weighted stylized funds, formed annually and based on the highest factor loadings, along with their corresponding benchmarks. It appears that the average monthly returns for a small-oriented stylized portfolio is marginally higher than the monthly average returns for a big-oriented stylized portfolio. For example, the average monthly return for small-growth stylized portfolio is 0.86% per month (i.e., 10.32% annually), while the average monthly return for big-value stylized portfolio is 0.61% per month (i.e., 7.32% annually). However, stylized portfolios except small-value have a lower yield return compared to its corresponding mimicking portfolios. The standard deviation for both stylized and mimicking portfolios is notably lower for growth-oriented portfolios than value-oriented portfolios. This suggests that value-oriented portfolios are more prone to extreme outcomes or are riskier than a growth-oriented portfolio. Thus, value-oriented investors seem to bear more risk with no compensation for value premium.

The most important comparison with respect to the descriptive statistics of the mimicking and stylized portfolios is that stylized portfolios with the exception of big-growth, have generated slightly lower dispersion than the mimicking portfolios indices and the market index "FTSE 100". One important conclusion from this result is that fund managers' selection skills have made the fund less risky compared to its benchmark. Although the distributions of both stylized and mimicking portfolios are fairly symmetrical, the distributions of small-oriented portfolios are highly leptokurtic with excess kurtosis ranging from 2.1 to 6.01. Moreover, the normality of returns is strongly rejected for all portfolios, as a Jarque-Bera test rejects the normality null hypothesis at the 5% significance level.

Table 5.2: Descriptive statistics of monthly return of 4 equally weighted stylized funds sorted by the highest RBSA factor weights on yearly basis, along with their corresponding benchmark.

	Stylized portfolios				*Mimicking Portfolios*				
	C4SG	***C4SV***	***C4BG***	***C4BV***	***SG***	***SV***	***BG***	***BV***	***Ftse100***
Observation	144	144	144	144	144	144	144	144	144
Mean (%)	0.86	0.82	0.71	0.61	1.02	0.71	0.93	0.67	0.66
Std Error (%)	4.02	4.43	3.60	4.39	4.56	5.63	3.19	4.51	4.11
t-statistcs	2.54*	2.1**	2.3**	1.47	2.6**	1.42	3.4**	1.42	1.7*
Skewness	-0.84	-0.6	-0.87	-0.74	-0.39	0.06	-0.51	-0.43	-0.46
Excess-Kurtosis	2.9	2.1	1.4	1.15	2.54	6.01	0.24	0.54	0.69
Jarque-Bera P-	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.02
Minimum (%)	-15.5	-16. 6	-12.4	-13.9	-14.6	-24.5	-9.03	-14.93	-11.30
Maximum (%)	13.60	14.70	9.30	9.30	18.33	27.72	7.81	12.19	11.80

Notes: At the end of each year from 2005 to 2016, the average loadings of four value weighted mimicking portfolios are computed using Sharpe (1992) regression: $R_{it} = \sum_{j=1}^{k} \omega_{jt} f_{jt} + e_{it}$. Thereafter, UK equity unit trusts are sorted into four stylized portfolios based on the highest weight exposure for the 12 months following year s. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

Table 5.3 reports the abnormal performance gross of management fees based on the time varying RBSA Sharpe (1992) approach as described in chapter 4.1.2. The result shows that the performances deteriorated with growth-oriented investment style regardless of the size factor. Big-growth stylized funds on average underperforms its benchmark (BG mimicking portfolio) by 0.21% per month (i.e., 2.52% annually). Small-growth stylized fund on average underperform its mimicking portfolio by 0.17% per month (i.e., 2.04% annually). On the other hand, funds tilted toward small-value investment have achieved on average 0.11% per month or 1.32% per annum higher return than small-value mimicking portfolio. However, the t-statistics associated with the average differences are statistically insignificant at any conventional level. The average abnormal performance relatively to the market index "FTSE100" showed that the performances are considerably positive for all stylized portfolios except for the big-value portfolio. Although, we are still unable to reject the null hypothesis that the relative performance is not different from zero at an acceptable level of significance, one would conclude that on average UK equity funds' managers do not possess any superior skills that allow them to beat their style specific benchmark and therefore to generate abnormal performance.

Table 5.4 reports risk-adjusted performance measures gross of management fees computed for 4 stylized portfolios and mimicking indices during the sample period of January 2005 to July 2017. It appears that the stylized portfolios in our sample slightly outperformed their corresponding mimicking portfolios except for the big-growth portfolio. The big-growth stylized portfolio underperformed its benchmark by 0.15 per month (i.e., 1.8% annually). In contrast, the outperformance varies

Table 5.3: Average abnormal performance of UK unit trusts based on continuous changing investment styles of Sharpe (1992) asset-classes model.

	Mimicking Portfolio			***FTSE 100***		
Series	***Average abnormal Returns***	***Std -Error***	***t-stat***	***Index Average Abnormal Returns***	***Std-Error***	***t-stat***
$\boldsymbol{AR_t^{SG}}$	-0.17%	2.1%	-1.0	0.18%	5.16%	0.59
$\boldsymbol{AR_t^{SV}}$	0.11%	2.59%	0.5	0.16%	5.44%	0.41
$\boldsymbol{AR_t^{BG}}$	-0.21%	2.12%	-1.23	0.04%	5.19%	0.24
$\boldsymbol{AR_t^{BV}}$	-0.06%	2.63%	-0.28	-0.04%	5.40%	-0.22

Notes: This table reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between stylized portfolios' return and its corresponding benchmark and FTSE100.($AR_t^{SG} = C4_t^{SG} - SG_t$ & $IAR_t^{SG} = C4_t^{SG} - ftse100_t$). Returns are expressed as percent per month.

between 0.01% per month (i.e., 0.12% annually) for the small-growth stylized fund and 0.28% (i.e., 3.36% annually) for the small-value stylized fund. The intercepts are, however, statistically insignificant at any conventional level, except for the small-value stylized portfolio. Therefore, investors with a preference for a small-value investment style are achieving higher risk-adjusted returns. However, the stylized portfolio performance is most likely to deteriorate with the introduction of transaction costs and management fees. The adjusted R-squared ranging from a low of 67 percent for C4BG to 81.2 percent for C4SV. The higher the percentage value of R-square, the more consistently fund managers are tracking their benchmark. Although the risk (volatility) associated with big-growth stylized funds is close to that of its benchmark (i.e., $\beta^{C4BG} = 0.91$), only 67 percent of C4BG return is explained by the variability of the mimicking portfolio. Thus, big-growth oriented funds are partially more active than its mimicking investment style, contain relatively little diversification within the big-growth asset class and are more likely to rotate across different investment styles.

Table 5.4: style regression performance of the UK unit trusts based on continuous changing investment styles(C4).

	Intercept	***p-Value***	***Beta***	***R^2***
C4SG	0.01%	0.9	0.78	79.5%
C4SV	0.28%	0.06	0.69	81. 2%
C4BG	-0.15%	0.41	0.91	67%
C4BV	0.08%	0.68	0.72	67.8%

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression: $C4_t^j - Rf_t = \hat{\alpha} + \hat{\beta}(j_t - Rf_t) + \hat{e}_t$. Where: $j = SG, SV, BG, and\ BV$, based on the period Jan 2005 to Jul 2017.

5.1.2. Result based on continuous changing style (C9)

Table 5.5 shows that the bulk of unit trusts are either invested in big-oriented shares or in a mixed investment style. Between 2005 and 2007 the funds are clustered around value-oriented shares and the market index (FTSE 350), while they shifted towards big-oriented shares between 2008 and 2011. Nevertheless, during the sovereign debt crisis, particularly between 2011-2013, the aggregate funds were more tilted towards the big-value investment style.

Table 5.6 shows the descriptive statistics of the nine equally weighted stylized funds, formed annually and based on factor loadings described in Table (5.1). The monthly average returns range from 1.01% for small-growth to 0.6% for the weak value investment style. Note that the pure/weak investment style refers to funds where factor exposure to one of the four mimicking portfolios is greater/lower than 0.5.

Table 5.5: Number of unit trusts in nine stylized portfolios based on the distribution of estimated factor loadings for unit trusts using 36-month rolling window.

Year	*ROL*	*TNF*	*NSG*	*NSV*	*NBG*	*NBV*	*NS*	*NB*	*NG*	*NV*	*NMix*
2005	37	121	6(5%)	12(10%)	8(7%)	1(1%)	7(6%)	36(30%)	0(0%)	27(22%)	24(20%)
2006	49	138	5(4%)	16(12%)	3(2%)	1(1%)	12(9%)	39(28%)	1(1%)	37(27%)	24(17%)
2007	61	154	8(5%)	16(10%)	6(4%)	0(0%)	26(17%)	17(11%)	2(1%)	56(36%)	23(15%)
2008	73	161	11(7%)	6(4%)	13(8%)	1(1%)	11(7%)	47(29%)	6(4%)	33(20%)	33(20%)
2009	85	167	2(1%)	9(5%)	60(36%)	1(1%)	8(5%)	37)22%)	7(4%)	5(3%)	38(23%)
2010	97	172	2(1%)	3(2%)	39(23%)	7(4%)	4(2%)	79(46%)	11(6%)	5(3%)	22(13%)
2011	109	175	4(2%)	3(2%)	28(16%)	31(18%)	2(1%)	74(42%)	9(5%)	4(2%)	20(11%)
2012	121	178	6(3%)	5(3%)	14(8%)	82(46%)	1(1%)	36(20%)	6(3%)	4(2%)	24(13%)
2013	133	179	8(4%)	9(5%)	5(3%)	62(35%)	4(2%)	23(13%)	16(9%)	8(4%)	44(25%)
2014	145	180	7(4%)	11(6%)	5(3%)	19(11%)	7(4%)	32(18%)	28(16%)	6(3%)	65(36%)
2015	157	184	19(10%	7(4%)	8(4%)	27(15%)	13(7%)	56(30%)	4(2%)	17(9%)	33(18%)
2016	169	181	29(16%)	2(1%)	93(51%)	1(1%)	1(1%)	15(8%)	0(0%)	27(15%)	13(7%(

Notes: The symbols used to denote the investment style imply the following: ROL = month, TNF= total number of studied funds, NSG = number of small-growth, NSV= number of small-value, NBG= number of big-growth, NBV= number of big-value, NS= number of small, NB= number of big, NG= number of growth-oriented, NV= number of value-oriented investment and NMix= number of mixed investments.

Generally, it appears that small oriented pure investment styles (i.e., C9SG & C9SV) have achieved the highest average monthly returns during the studied period. For example, the monthly average return for the small-growth stylized portfolio is 1.01% (i.e., 12.12% annually) statistically significant at 1% level of significance. The dispersion of returns is significantly higher for small and value stylized portfolio. For example, the small-value stylized portfolio has a standard deviation of 4.83% per month, in comparison to 3.43% for the big-growth stylized portfolio. Moreover, the distribution of returns across all stylized portfolios are fairly symmetrical and highly leptokurtic, indicating heavier tails/outliers than expected from normal distribution.

Table 5.7 reports the abnormal performance gross of management fees calculated by taking the monthly return difference between the stylized portfolios' return and either its corresponding mimicking portfolio or the FTSE100. It appears that the performance varies across the stylized funds. The average performance with respect to mimicking portfolios ranges from 0.16% per month (1.92% annually) for big-value funds to -0.27% per month (-3.24 % annually) for the weak growth stylized portfolio. The underperformance documented in the stylized funds is mainly attributed to the growth effect, whereby growth-oriented funds considerably underperformed its corresponding benchmarks. However, the t-statistics associated with the average differences are statistically insignificant at any conventional level, except for the weak growth stylised portfolio. In contrast, the average abnormal performance relative to the market index "FTSE100" reveals that the performances are positive for all stylized portfolios. It is worth noting that the pure investment style except big-growth have achieved the highest performance across the nine stylized portfolios. In particular, small-growth and small-value have yields of 0.35% and 0.20% per month respectively. The standard deviation of performance is significantly different to that found from the mimicking portfolios. However, the t-statistics estimates indicate that the performance is statistically insignificant at any conventional level.

Table 5.6: Descriptive statistics of monthly return of 9 equally weighted funds' return sorted by the first and second highest RBSA factor weights on yearly basis.

Stylized portfolios	***C9SG***	***C9SV***	***C9BG***	***C9BV***	***C9S***	***C9G***	***C9V***	***C9B***	***C9MIX***
Observation	144	144	144	132	144	144	120	144	144
Mean	1.01%	0.86%	0.73%	0.80%	0.74%	0.70%	0.60%	0.68%	0.74%
Std Error	4.24%	4.83%	3.43%	4.90%	4.44%	3.69%	4.31%	3.72%	3.87%
t-statistic	2.87**	2.11*	2.54*	1.80	2.0*	2.28*	1.52	2.21*	2.28*
Skewness	-1.15	-0.05	-0.85	-1.27	-0.74	-0.95	-0.70	-0.78	-0.50
Excess-Kurtosis	4.20	1.70	1.16	4.70	3.40	1.95	1.30	1.15	0.80
Jarque-Bera P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Minimum (%)	-18.3%	-15.3%	-11.0%	-22.5%	-18.2%	-13.3%	-10.7%	-12.3%	-14.5%
Maximum (%)	12.7%	18.2%	9.1%	11.8%	17.5%	9.2%	9.4%	9.3%	11.2%

Notes: At the end of each year from 2005 to 2017, UK equity unit trusts are sorted into 9 stylized portfolios based on the highest and the second highest loadings of four value weighted mimicking portfolios. Where the factor weights are estimated using Sharpe (1992) regression: $R_{it} = \sum_{j=1}^{k} \omega_{jt} f_{jt} + e_{it}$. The symbols used to denote the investment style imply the following: C9SG = small-growth, C9SV=small-value, C9BG=big-growth, and C9BV=big-value, C9S=small-oriented, C9G=Growth-oriented, C9V=value-oriented, C9b= big oriented and C9BL=Balanced investment style.** Indicates significance at the 1 percent level.* Indicates significance at the 5 percent level.

Table 5.7: Average abnormal performance of 9 Stylized portfolio based on continuous changing investment styles of Sharpe (1992) asset-classes model.

	Mimicking Portfolio			***FTSE 100***		
Series	***Average Abnormal Returns***	***Std Error***	***t-stat***	***Index Average Abnormal Returns***	***Std Error***	***t-stat***
AR_t^{C9SG}	-0.01%	2.16%	-0.05	0.35%	5.09%	0.40
AR_t^{C9SV}	0.15%	2.48%	0.72	0.20%	5.87%	0.15
AR_t^{C9BG}	-0.20%	2.02%	-1.18	0.07%	5.11%	0.31
AR_t^{C9BV}	0.16%	3.63%	0.51	0.16%	5.92%	0.17
AR_t^{C9S}	-0.13%	1.97%	-0.77	0.08%	5.45%	0.09
AR_t^{C9G}	-0.27%	1.57%	2.09*	0.04%	5.08%	0.1
AR_t^{C9V}	0.05%	2.04%	0.24	0.05%	5.75%	0.1
AR_t^{C9B}	-0.12%	1.73%	-0.81	0.02%	5.28%	0.05
AR_t^{C9BAL}	0.03%	5.24%	0.06	0.08%	5.27%	0.17

Notes: This table reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between stylized portfolios' return and its corresponding benchmark and FTSE100.($AR_t^{C9SG} = C9_t^{SG} - SG_t$ & $IAR_t^{C9SG} = C9_t^{SG} - ftse100_t$). Returns are expressed as percent per month. * Indicates significance at the 5 percent level.

Table 5.8 shows that the intercepts from the regression of equation 4.5 across different investment styles. Generally, the abnormal performance is more pronounced in the value-oriented stylized portfolios, while underperformance is more likely to be present in the growth-oriented stylized portfolio. The performance ranges between -0.38% per month (-4.56 per year) for weak-growth to 0.27% per month (3.24% per year) for the small-value stylized portfolio. The intercepts are, however, mostly statistically insignificant at any conventional level, except for the weak growth investment style. The weak growth stylized portfolio achieved -0.38% per month under its mimicking portfolio, while imitating 99% of its risk ($\beta^{C9G} = 0.99$). Moreover, 79.9% of the variation in the weak growth stylized portfolio can be explained by the variation of its mimicking portfolio. By comparing these results with those from the C4 portfolios formation method in Table 5.4, we can see a similar performance trend (i.e., underperformance is more pronounced for growth-oriented stylized portfolios. The C9 beta coefficients show that there is a slightly higher correlation between the C9 pure stylized portfolios and its corresponding benchmarks than those observed under the C4 approach. The adjusted R-squared is marginally lower for C9 pure stylized portfolios than most stylized portfolios. However, this might be simply due to a low number of funds in the C9 pure stylized portfolios (see Table 5.5).

Table 5.8: Style regression performance of the UK unit trusts based on continuous changing investment styles(C9).

	Intercept	***p-Value***	***Beta***	***R^2***
C9SG	0.14%	0.41	0.82	78.9%
C9SV	0.27%	0.10	0.77	81.3%
C9BG	-0.10%	0.52	0.88	67.1%
C9BV	0.28%	0.35	0.76	51.2%
C9S	-0.002%	0.98	0.82	85.6%
C9G	-0.38 %	0.03	0.99	79.9%
C9V	0.17%	0.24	0.69	76.6%
C9B	-0.07%	0.60	0.94	76.3%
C9BAL	0.12%	0.3	0.88	88.2%

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression: $C9_t^j - Rf_t = \hat{\alpha} + \hat{\beta}(j_t - Rf_t) + \hat{e}_t$. Where: $j = SG, SV, BG, BV, S, B, G, V and\ Bal$, based on the period Jan 2005 to Jul 2017.

5.2. Dominant Style

5.2.1. Result based on Dominant Style (D4)

Similar to the results obtained from C4, Table 5.9 shows that unit trusts that are tilted towards small stocks performed, on average, slightly better than the big-oriented funds. The average performance ranges between 0.54% to 0.76% per month for the big-value and small-value stylized portfolios respectively. The standard deviation is marginally lower for big-oriented funds. Furthermore, the distributions of the stylized portfolios formed using the D4 approach are fairly symmetrical, but highly leptokurtic. Row 4 of Table 5.9 demonstrates that given the number of funds in each stylized portfolio, it appears that more than 80 percent of the studied funds are tilted towards the big-oriented investment style.

Table 5.10 presents the performance of the four stylized portfolios relative to their mimicking portfolios and the FTSE100. The results are found to be similar to the result obtained from C4 approach in Table 5.3. The gross performance deteriorates with the growth-oriented investment style regardless of the size effect. For example, small-growth and big-growth stylized portfolios underperformed their mimicking portfolios by -0.29% and -0.21% per month, respectively. However, the t-statistics associated with the average differences across the four stylized portfolios are statistically insignificant at any conventional level. In contrast, the average abnormal performance relative to the market index “FTSE100” shows that the performances are considerably positive for all stylized portfolio except for the big-value portfolio.

Table 5.9: Descriptive statistics of monthly return of 4 equally-weighted funds' return sorted by the highest RBSA factor weights for the whole period. along with their corresponding benchmark.

Stylized portfolios					***Mimicking Portfolios***				
	D4SG	***D4SV***	***D4BG***	***D4BV***	***SG***	***SV***	***BG***	***BV***	***Ftse100***
TNF	24	11	77	69	NA	NA	NA	NA	NA
Mean (%)	0.73	0.76	0.71	0.54	1.02	0.71	0.93	0.67	0.66
Std Error (%)	4.08	4.13	3.65	3.88	4.56	5.63	3.19	4.51	4.11
t-statistic	2.15*	2.19*	2.31*	1.69	2.6**	1.42	3.4**	1.42	1.7*
Skewness	-0.78	-0.8	-0.81	-1.05	-0.39	0.06	-0.51	-0.43	-0.46
Excess-Kurtosis	2.02	2.1	1.2	2.72	2.54	6.01	0.24	0.54	0.69
Jarque-Bera P-	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.02
Minimum (%)	-15.2	-16. 0	-12.3	-16.7	-14.6	-24.5	-9.03	-14.93	-11.30
Maximum (%)	11.5	13.2	9.3	8.8	18.33	27.72	7.81	12.19	-11.80

Notes: At the end of each year from 2005 to 2017, the average loadings of four value weighted mimicking portfolios are computed using Sharpe (1992) regression: $R_{it} = \sum_{j=1}^{k} \omega_{jt} f_{jt} + e_{it}$. Thereafter, UK equity unit trusts are sorted into four stylized portfolios based on the average highest weight exposure for the whole studied period. The symbols used to denote the investment style imply the following: TNF= total number of funds, SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

Nevertheless, we are still unable to reject the null hypothesis that the relative performance is no different than zero at an acceptable confidence level. However, it is worth noting that the performance of our stylized portfolios has generated a lower dispersion relative to the mimicking portfolios than the market index (FTSE 100).

Similarly, from Table 5.11, the regression-based performance reveals that the stylized portfolios' performance declined in line with growth-oriented portfolios regardless of the size effect. The big-growth stylized portfolio on average underperformed its benchmark by -0.16% per month (-1.92% annually), as opposed to 0.23% per month (2.76 % annually) for small-value stylized portfolio. However, Alphas are statistically insignificant at an accepted level. The regression coefficients showed that growth-oriented stylised portfolios have higher betas than value-oriented stylized portfolios and, hence, track their designated mimicking portfolios closely. Hence, the return of value-oriented stylized portfolios either contains residual risk which are not correlated with the mimicking portfolios or fund managers were able to beat their benchmark and achieve lower risk compared to their mimicking portfolios.

By comparing the results obtained from the C4 and D4 approach, we can see that value-oriented fund managers on average do better than growth-oriented managers on a style-adjusted basis. Whether the performance is measured relative to style mimicking portfolio or single factor regression model, the intercepts are statistically not different from zero. Thus, investors can be indifferent when considering between investing in a style mimicking index or any other actively managed fund. However, the performance of an actively managed fund is most likely to deteriorate with the introduction of transaction costs and management fees, although under the single factor model, there is some evidence that, on average, small-value stylized portfolios are able to generate statistically superior performance over its corresponding benchmark.

Table 5.10: Average abnormal performance of UK unit trusts based on dominant style for the whole studied period.

		Mimicking Portfolio			*FTSE 100*		
Series	***Observation***	***Average***	***Std Error***	***t-stat***	***Index Average***	***Std Error***	***t-stat***
$\boldsymbol{AR_t^{SG}}$	144	-0.29%	2.4%	-1.45	0.06%	5.54%	0.15
$\boldsymbol{AR_t^{SV}}$	144	0.04%	2.75%	0.18	0.09%	5.35%	0.20
$\boldsymbol{AR_t^{BG}}$	144	-0.21%	2.21%	-1.19	0.04%	5.24%	0.10
$\boldsymbol{AR_t^{BV}}$	144	-0.02%	2.68%	-0.57	-0.11%	5.33%	-0.25

Notes: This table reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between stylized portfolios' return and its corresponding benchmark and FTSE100.($AR_t^{SG} = D4_t^{SG} - SG_t$ & $IAR_t^{SG} = D4_t^{SG} - ftse100_t$). Returns are expressed as percent per month.

Table 5.11: Style regression performance of the UK unit trusts based on Dominant investment styles(D4).

	Intercept	***p-Value***	***Beta***	***R^2***
D4SG	-0.08%	0.64	0.76	73.1%
D4SV	0.23%	0.13	0.65	79. 1%
D4BG	-0.16%	0.40	0.92	65.5%
D4BV	0.02%	0.88	0.69	65.6%

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression: $D4_t^j - Rf_t = \hat{\alpha} + \hat{\beta}(j_t - Rf_t) + \hat{e}_t$. Where: $j = SG, SV, BG, and\ BV$, based on the period Jan 2005 to Jul 2017.

5.3. Ethical Unit Trusts

In this section we turn our attention to ethical funds investment style, particularly our aim is to test whether ethical fund managers invest differently than their conventional counterparts and if ethical fund investors pay a higher price for their ethical consideration. The objective of this section is twofold. First, if ethical funds are positioned to match a particular style, then a style benchmark is ordinarily used against which performance might be evaluated. The second purpose is with the intention to provide significant complementary evidence on ethical mutual fund performance. Employing the Sharpe (1992) asset-classes factor model, we capture individual ethical funds exposure to four asset classes. Then, we identify an appropriate benchmark to evaluate the performance of ethical funds. Similar procedures to those already described earlier in this chapter are applied to form stylized portfolios of ethical funds. However, since the number of UK-equity ethical funds in our sample is low (32 ethical funds), we are only able to construct four portfolios of stylized unit trusts based on the highest factor exposure produced by RBSA regression (i.e., C4&D4). Thus, we examine whether there is any difference in performance across investment styles. Finally, we compare the performance of stylized ethical portfolios against their conventional peers, using investment style as the matching criteria. We are therefore able to account for the possible return differences between ethical and conventional funds with respect to their investment style.

5.3.1. Ethical Continuous Changing Style (EC4)

Table 5.12 quantifies the number of funds in four stylized portfolios (EC4) based on the distribution of the highest estimated factor loading using a 36-month rolling window. It is clear that before 2011 ethical unit trusts were tilted towards growth stocks; more than two thirds were invested in either big-growth or small-growth stocks. However, funds' exposure to small-growth stocks has declined steadily during the studied period, whilst investments tilted towards big-value stocks have increased gradually. This result contradicts the finding of Gregory and Whittaker (2007), who observed that UK ethical funds exhibit significant exposure to small firms.

Table 5.12: Number of Ethical unit trusts grouped into four stylized portfolios based on the highest estimated factor loadings using 36-month rolling window.

Year	***Rol***	***NF***	***SG***	***SV***	***BG***	***BV***
2005	37	19	8	2	7	2
2006	49	19	9	2	6	2
2007	61	19	7	3	6	3
2008	73	20	6	4	7	3
2009	85	21	6	5	7	3
2010	97	22	5	5	8	4
2011	109	23	5	4	10	4
2012	121	22	4	3	9	6
2013	133	20	3	1	8	8
2014	145	19	3	1	8	7
2015	157	20	3	1	8	8
2016	169	18	2	0	16	0

The symbols used to denote the investment style imply the following: ROL = time, TNF= total number of studied funds, NSG = number of small-growth, NSV= number of small-value, NBG= number of big-growth, and NBV= number of big-value investment style.

Table 5.13 shows the descriptive statistics of four equally weighted ethical funds' annually formed returns, based on the highest factor loadings. It is clear that the returns deteriorated with the value effect, whereby growth-oriented ethical funds have on average generated approximately 0.6% per month (7.2% per year), almost 0.2% per month higher than value-oriented ethical funds. The dispersion of returns for growth-oriented stylized portfolios is also slightly lower. The t-statistic estimates show that the mean returns are only significantly different from zero as funds approach the growth end of the style spectrum. While the distributions of the stylized portfolios are fairly symmetrical and highly leptokurtic. The average monthly returns for ethical stylized portfolios are considerably lower compared to their mimicking benchmarks. However, ethical funds were able to achieve lower deviation in returns with the exception of the big-growth stylized portfolio.

Table 5.14 reports the monthly average performance of four stylized ethical portfolios relative to their corresponding mimicking portfolios and the market index (FTSE4GOOD) gross of management fees. It appears that stylized ethical portfolios have underperformed their corresponding mimicking portfolios, except for the small-value ethical portfolio. The performance varies between -0.41% per month for small-growth funds and 0.09% for small-value ethical funds. However, the t-statistics are significant only for growth-oriented portfolios at the 5% level for small-growth and the 10% level for the big-growth stylized portfolio. These results indicate that ethical funds tilted towards a growth-oriented investment style are more likely to underperform their style benchmark. On the other hand, the average performances relative to the Ftse4good index are mostly negative, except for the small-growth stylized portfolio. Nonetheless, t-statistics show that, across all of the stylized portfolios, the deviation in performance is not significantly different from zero. The deviation in performance is however significantly higher than those observed from the mimicking portfolios. Hence, the returns of the stylized portfolios are more consistent with the style mimicking portfolios than the market index (FTSE4GOOD).

Table 5.15 reports the performance estimates for equally weighted stylized ethical funds using the single factor model. It appears that the stylized portfolios in our sample generally underperformed their corresponding benchmarks except for the small-value portfolio. The average alpha estimates range between -0.2% and 0.1% per month for big-growth and small-value, respectively. However, the intercepts are statistically insignificant at any accepted level. The coefficient estimates vary from 0.6 for small-value and 0.92 for big-growth stylized portfolio. For example, the big-growth portfolio tracks 92% of its style mimicking portfolio. The adjusted R-squared ranges are between 63% and 75.3%, thus the selected mimicking portfolio explains 63% to 75% of the variance in a typical unit trust. Furthermore, the adjusted R-squared value is slightly lower for big-oriented funds, which might suggest that big-oriented ethical fund managers pursue active management.

.

Table 5.13: Descriptive statistics of monthly return of 4 equally weighted ethical funds' return sorted by the highest RBSA factor weights on yearly basis, along with their corresponding benchmark.

	Stylized portfolios				***Mimicking Portfolios***				
	EC4SG	***EC4SV***	***EC4BG***	***EC4BV***	***SG***	***SV***	***BG***	***BV***	***Ftse4good***
Observation	144	108	144	132	144	144	144	144	144
Mean (%)	0.66	0.4	0.62	0.46	1.03	0.72	0.97	0.56	0.64
Std Error (%)	4.01	4.17	3.71	4.38	4.53	5.79	3.21	4.55	4.20
t-statistic	1.85*	0.99	1.94*	1.22	2.6**	1.42	3.4**	1.42	1.77*
Skewness	-1.11	-1.03	-0.93	-1.19	-0.39	0.06	-0.51	-0.43	-0.43
Excess-Kurtosis	2.53	2.7	2.2	3.3	2.54	6.01	0.24	0.54	0.78
Jarque-Bera P-	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.02
Minimum (%)	-17.1	-17. 8	-14.6	-20.3	-14.6	-24.5	-9.03	-14.9	-11.30
Maximum (%)	8.4	11	10.33	11.08	18.33	27.7	7.81	12.19	-11.80

Notes: At the end of each year from 2005 to 2016, the average loadings of four value weighted mimicking portfolios are computed using Sharpe (1992) regression: $R_{it} = \sum_{j=1}^{k} \omega_{jt} f_{jt} + e_{it}$. Thereafter, UK ethical unit trusts are sorted into four stylized portfolios based on the highest weight exposure for the 12 months following year s. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

Table 5.14: Average abnormal performance of UK Ethical unit trusts based on continuous changing investment styles of Sharpe (1992) asset-classes model.

		Mimicking Portfolio			FTSE4good		
Series	***Observation***	***Average Abnormal Returns***	***Std -Error***	***t-stat***	***Index Average Abnormal Returns***	***Std- Error***	***t-stat***
$\boldsymbol{AR_t^{ESG}}$	144	-0.41%	2.57%	-1.92*	0.01%	5.41%	0.03
$\boldsymbol{AR_t^{ESV}}$	108	0.09%	3.18%	0.30	-0.05%	5.5%	-0.09
$\boldsymbol{AR_t^{EBG}}$	144	-0.03%	2.26%	-1.75	-0.01%	5.31%	-0.04
$\boldsymbol{AR_t^{EBV}}$	132	-0.09%	2.81%	-0.38	-0.17%	5.57%	-0.37

Notes: This table reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between ethical stylized portfolios' return and its corresponding benchmark and FTSE4good.($AR_t^{ESG} = EC4_t^{SG} - SG_t$ & $IAR_t^{ESG} = EC4_t^{SG} - ftse4good_t$). Returns are expressed as percent per month. * Indicates significance at the 5 percent level.

Table 5.15: Style regression performance of the UK ethical unit trusts based on continuous changing investment styles(C4).

	Intercept	*p-Value*	*Beta*	*R^2*
EC4SG	-0.1%	0.18	0.75	68.4%
EC4SV	0.13%	0.48	0.60	75.3%
EC4BG	-0.2%	0.19	0.92	63.6%
EC4BV	-0.003%	0.98	0.76	63%

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression: $EC4_t^j = \hat{\alpha} + \hat{\beta} j_t + \hat{e}_t$. Where: $j = ESG, ESV, EBG, and\ EBV$, based on the period Jan 2005 to Jul 2017.

Next, we turn our attention to examining the performance of UK ethical unit trusts against their conventional peers. To account for the possible return differences between ethical and conventional funds, we compare the performance of ethical funds with a matched sample of conventional funds using the investment style as the matching criteria. Thus, each stylized ethical portfolio is matched with the appropriate stylized conventional portfolio (i.e., EC4SG is matched with C4SG).

Table 5.16 panel A reports the result based upon the relative performance of four stylized ethical portfolios and their corresponding stylized conventional portfolio. Over the entire sample period, the average stylized ethical portfolios earned a significantly lower average monthly return than its conventional peers. The underperformance is more pronounced for small-oriented stylized portfolios, for example, $\boldsymbol{C4ESV}$ on average underperformed relative to the $\boldsymbol{C4SV}$ stylized portfolio by 0.5% per month (6% per annum), while $\boldsymbol{C4EBV}$ on average underperformed the $\boldsymbol{C4BV}$ stylized portfolio by 0.24% per month (2.88% per annum). Across all stylized portfolios, the difference in performances is statistically significant at a conventional level, except for the big-value portfolio. Furthermore, by comparing the standard deviation of stylized ethical funds in Table 5.9 and conventional funds in Table 5.2, it can be seen that ethical stylized portfolios are neither more nor less risky than their conventional peers. It is worth noting that each ethical stylized portfolio contains fewer funds compared to its conventional counterpart, and hence a less diversified portfolio.

In panel B of Table 5.16 we report the results from a zero net investment strategy that can be achieved by long the stylized ethical portfolios and short the stylized conventional portfolios. Similar to the previous result, the average stylized ethical portfolios yield significantly lower average returns than their conventional counterpart. However, the disappointing performance is more pronounced among funds managers who have a preference for small market cap stocks. The intercepts are statically significant at the 1% level for small-growth and small-value stylized portfolios, respectively. The beta coefficients are negative, which indicate that stylized ethical portfolios have lower style exposure than stylized conventional portfolios. The results are different from those of Bauer et al. (2005) and Gregory and Whittaker (2007) who report insignificant performance on a risk and style adjusted basis. Our findings are more with consistent Gregory et al. (1997) who suggest that ethical funds' performance is inferior to that of conventional funds after controlling for risk and style characteristics. In Particular, ethical funds that are tilted toward a small-oriented investment style are more likely to underperform their conventional peers.

Table 5.16: Average abnormal performance of UK ethical vs conventional unit trusts based on continuous changing investment styles.

		Panel A			***Panel B***			
		Relative performance			***Regression-based performance***			
Series	***Observation***	***Average Abnormal Returns***	***Std -Error***	***t-stat***	***Intercept***	***p-Value***	***Beta***	***R^2***
$\boldsymbol{AR_t^{ESG}}$	144	-0.24%	1.18%	-2.49**	-0.20%	0.04	-0.05	8.3%
$\boldsymbol{AR_t^{ESV}}$	144	-0.5%	2.22%	-2.82**	-0.17%	0.00	-0.17	20.2%
$\boldsymbol{AR_t^{EBG}}$	144	-0.13%	0.75%	-2.20*	-0.11%	0.08	-0.03	5.3%
$\boldsymbol{AR_t^{EBV}}$	144	-0.18%	1.9%	-1.15	-0.01%	0.25	-0.01	1.02%

Notes: Panel A reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between ethical stylized portfolios' return and its corresponding conventional stylized portfolios' return.($AR_t^{SG} = EC4_t^{SG} - C4_t^{SG}$). Panel B reports the coefficients from calendar time regressions for investment strategy that can be achieved by long ethical and short conventional stylized portfolios. Regression equation is given as: $EC4_t^j - C4_t^j = \hat{\alpha} + \hat{\beta}(j_t - Rf_t) + \hat{e}_t$. Where: $j = SG, SV, BG, and\ BV$, based on the period Jan 2005 to Jul 2017. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

5.3.2. Dominant Style (ED4)

In this section we report the results obtained by applying a dominant style for each ethical fund throughout the sample period. This approach allows us to evaluate the performance of ethical funds using a static investment style (ED4) as opposed to a time-varying investment style (EC4).

Table 5.17 shows similar descriptive statistics to those observed under the time-varying investment style. Generally, the average performance of ethical portfolios formed using the D4 approach is much lower compared to that of stylized mimicking portfolios. The standard deviation is significantly lower for ethical portfolios than for the stylized mimicking portfolios.

Table 5.18 shows that the average performance of the analysed ethical funds relative to their mimicking portfolios is negative, but only significant for the small-growth stylized ethical portfolio. The average performance relative to the FTSE4good index is statistically not different from zero across the stylized ethical portfolios. This might suggest that ethical funds managers track the FTSE4GOOD more closely than their style benchmark.

Likewise, the results from a basic Jensen's alpha model in Table 5.19 reveal that the underperformance is statistically not different from zero for all stylized ethical portfolios. The style exposure (beta coefficients) is significant, and the big-growth stylized ethical portfolio has a significantly greater exposure, close to one.

Table 5.20 compares the performance of stylized ethical portfolios with a matched sample of conventional portfolios formed on a dominant investment style basis. The results do not show any support for ethical funds' underperformance compared to its conventional peers. This result contradicts the poor performance that is seen in the ethical funds' via the continuous changing style funds formation. We can therefore conclude that the performance of ethical funds is influenced by whether a dominant or continuous changing style is used in the portfolio formation method.

Table 5.17: Descriptive statistics of monthly return of 4 equally weighted ethical funds' return sorted by the highest RBSA factor weights for the whole period along with their corresponding benchmark.

	Stylized portfolios				***Mimicking Portfolios***				
	ED4SG	***ED4SV***	***ED4BG***	***ED4BV***	***SG***	***SV***	***BG***	***BV***	***Ftse4good***
Observation	144	144	144	144	144	144	144	144	144
Mean (%)	0.61	0.58	0.56	0.46	1.03	0.72	0.97	0.56	0.64
Std Error (%)	3.09	3.7	3.7	4.01	4.53	5.79	3.21	4.55	4.20
t-statistic	2.03*	1.92*	1.88*	1.38	2.6**	1.42	3.4**	1.42	1.77*
Skewness	-1.20	-0.99	-0.98	-1.19	-0.39	0.06	-0.51	-0.43	-0.43
Excess-Kurtosis	3.19	2.63	2.77	4.15	2.54	6.01	0.24	0.54	0.78
Jarque-Bera P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.02
Minimum (%)	-17.9	-16. 9	-16.5	-20.2	-14.6	-24.5	-9.03	-14.9	-11.30
Maximum (%)	9.1	10.8	11.35	11.58	18.33	27.7	7.81	12.19	-11.80

Notes: At the end of each year from 2005 to 2016, the average loadings of four value weighted mimicking portfolios are computed using Sharpe (1992) regression: $R_{it} = \sum_{j=1}^{k} \omega_{jt} f_{jt} + e_{it}$. Thereafter, UK ethical unit trusts are sorted into four stylized portfolios based on the highest weight exposure for the whole period. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

Table 5.18: Average abnormal performance of UK Ethical unit trusts based on dominant investment styles.

		Mimicking Portfolio			***FTSE4good***		
Series	***Observation***	***Average Abnormal Returns***	***Std -Error***	***t-stat***	***Index Average Abnormal Returns***	***Std- Error***	***t-stat***
$\boldsymbol{AR_t^{ESG}}$	144	-0.39%	2.4%	-1.93*	-0.04%	5.20%	-0.1
$\boldsymbol{AR_t^{ESV}}$	144	-0.10%	3.2%	-0.38	-0.07%	5.26%	-0.17
$\boldsymbol{AR_t^{EBG}}$	144	-0.35%	2.3%	-1.80	-0.09%	5.17%	-0.22
$\boldsymbol{AR_t^{EBV}}$	144	-0.1%	2.7%	-0.84	-0.19%	5.5%	-0.41

Notes: This table reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between ethical stylized portfolios' return and its corresponding benchmark and FTSE4good.($AR_t^{ESG} = EC4_t^{SG} - SG_t$ & $IAR_t^{ESG} = EC4_t^{SG} - ftse4good_t$). Returns are expressed as percent per month. * Indicates significance at the 5 percent level.

Table 5.19: Style regression performance of the UK ethical unit trusts based on dominant investment styles(D4).

	Intercept	*p-Value*	*Beta*	*R^2*
ED4SG	-0.1%	0.57	0.73	71.9%
ED4SV	0.19%	0.25	0.56	71.1%
ED4BG	-0.28%	0.20	0.92	62.2%
ED4BV	-0.01%	0.93	0.73	65.3%

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression: $ED4_t^j = \hat{\alpha} + \hat{\beta} j_t + \hat{e}_t$. Where: $j = ESG, ESV, EBG, and\ EBV$, based on the period Jan 2005 to Jul 2017.

Table 5.20: Average abnormal performance of UK ethical vs conventional unit trusts based on dominant investment styles.

		Panel A			***Panel B***			
		Relative performance			***Regression-based performance***			
Series	***Observation***	***Average Abnormal Returns***	***Std -Error***	***t-stat***	***Intercept***	***p-Value***	***Beta***	***R^2***
$\boldsymbol{AR_t^{ESG}}$	144	-0.1%	0.8%	-1.42	-0.08%	0.26	-0.02	1.3%
$\boldsymbol{AR_t^{ESV}}$	144	-0.01%	1.3%	-1.31	-0.01%	0.33	-0.08	12.2%
$\boldsymbol{AR_t^{EBG}}$	144	-0.13%	0.86%	-1.83	-0.13%	0.11	-0.00	2.3%
$\boldsymbol{AR_t^{EBV}}$	144	-0.05%	0.86%	-0.78	-0.07%	0.28	0.04	5.3 %

Notes: Panel A reports the average abnormal returns for each stylized portfolio by taking the monthly return difference between ethical stylized portfolios' return and its corresponding conventional stylized portfolios' return.($AR_t^{SG} = ED4_t^{SG} - D4_t^{SG}$). Panel B reports the coefficients from calendar time regressions on equally weighted portfolios of ethical and conventional stylized portfolios. Regression equation is given as: $ED4_t^j - D4_t^j = \hat{\alpha} + \hat{\beta}(j_t - Rf_t) + \hat{e}_t$. Where: $j = SG, SV, BG, and\ BV$, based on the period Jan 2005 to Jul 2017. ** Indicates significance at the 1 percent level. * Indicates significance at the 5 percent level.

5.4. Implementing Return-Based Style Analysis (RBSA) for individual fund.

In order to demonstrate how RBSA is applied in practice, we analysed two individual UK equity funds. We identified the funds' investment style by comparing the funds overall asset allocation to popular style mimicking portfolios. The return of each style mimicking portfolio was obtained from Exeter University's Centre for Finance and Investment (Gregory et al., 2013) and represents a passive investment strategy. Figure 5.1 shows the variation of the mimicking portfolios' return and market indices on a yearly basis between 2005 and 2017. We can see that the small-value mimicking portfolio have the highest variability in returns, while unsurprisingly FTSE indices have moderate variability compared to style mimicking portfolios.

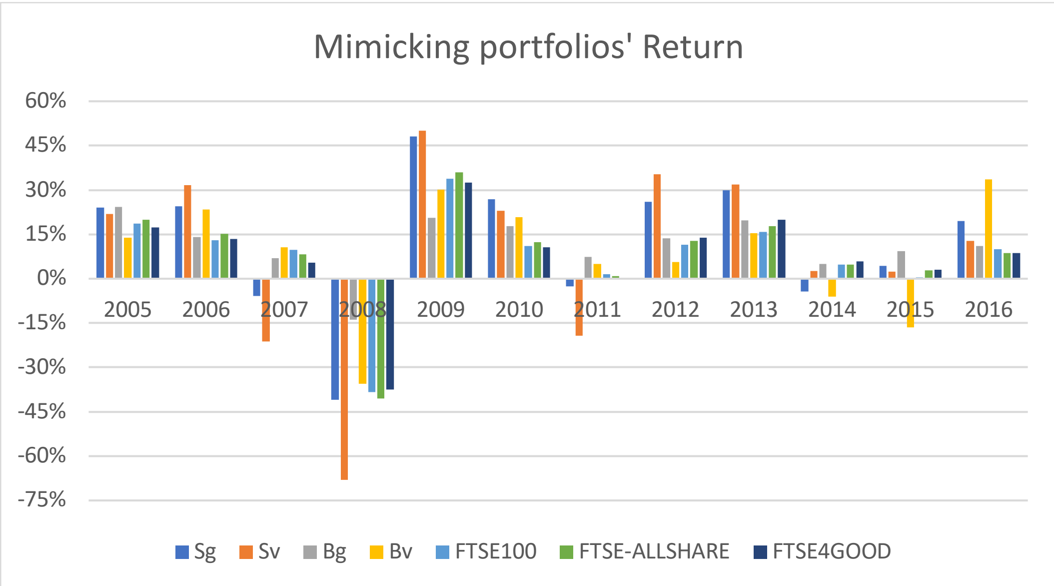


Figure 5.1: Annual mimicking portfolios' return between 2005 and 2017.

Style mimicking portfolios capture the size and growth-value dimensions that might explain the variation in the fund's returns. Thus, if fund managers' investment style is consistent with these dimensions, then funds' returns should mimic their style mimicking portfolios more closely than the market index. Accordingly, we can use the weights (exposures) to style mimicking portfolios to generate a fund customised benchmark. Then we can compare the funds realized return to the customized benchmark over the subsequent period. To illustrate the application of the RSBA, we investigate the following funds' investment style as an example.

5.4.1. Halifax UK Growth

The Halifax fund is classified as a large growth fund by Morningstar, with assets under management of £4.3 billion as of December 2018 (Morningstar 2018, a). Our first approach is a continuous changing style. We used the asset-classes factor model to compute the fund realized returns' exposure to the four asset classes over 36-month rolling windows. The factor exposure is allowed to vary on a yearly basis, with each set having 24 months in common with its predecessor. Figure 5.2 illustrates the trend of asset allocation in the Halifax UK growth investment portfolio. It shows that the biggest proportion of funds were allocated for big-oriented stocks, and ranges between 70-80 % of total asset holdings across the whole period. The proportion of funds that went to small-oriented stocks remained relatively constant, between 20-30 %, in the same period. However, the proportion of big-value investment steadily increased throughout the period. For example, fund managers increased their big-value stock holdings from 20% in 2007 to a high of 60% by 2012.

Our second approach is the dominant style. Figure 5.3 presents the result based on the average changing styles over the period covered between 2005 and 2017. The bar chart implies that the Halifax fund has the following average exposure: 13%, 9%, 34%, and 43% for small-growth, small-value, big-growth, and big-value, respectively.

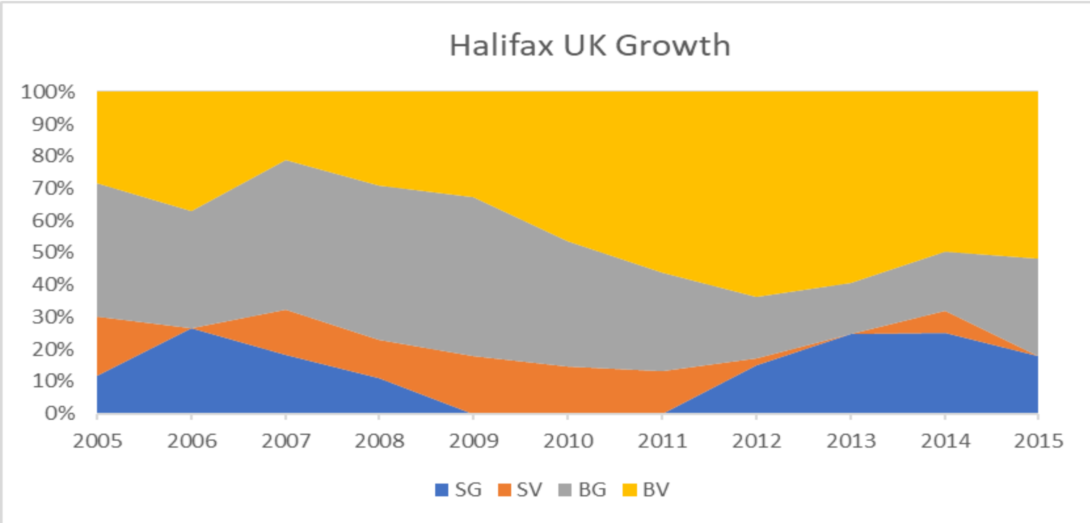


Figure 5.2: The estimated fund’s changing styles over the period covered.

To generate a return similar to the Halifax fund’s return, one would invest 13% in small-growth, 9% in small-value, 34% in big-growth, and 43% in a big-value style mimicking portfolio. Our result is consistent with the Morningstar classification (style box), which indicate that the Halifax fund is tilted toward big-oriented stocks, whereby 77% of the funds’ return is attributed to the big size effect.

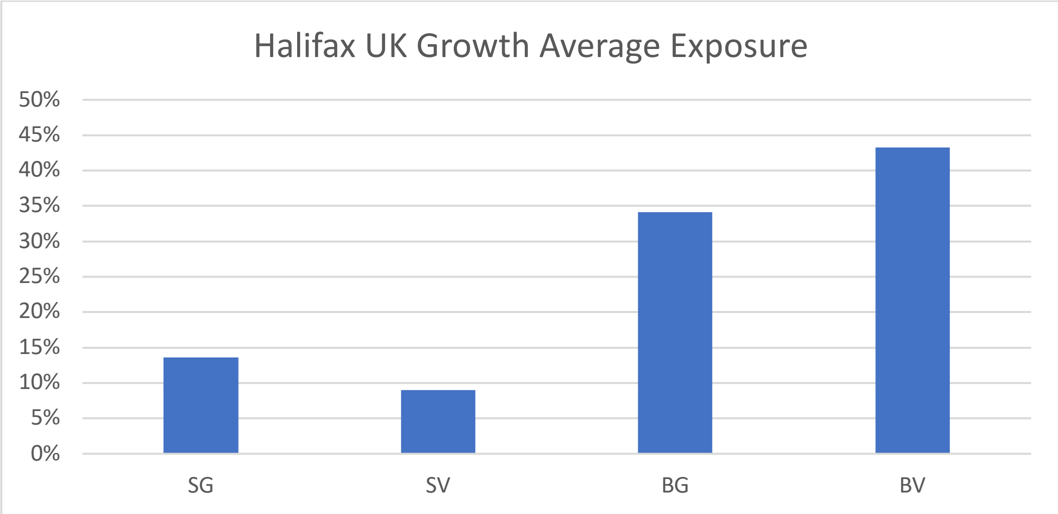


Figure 5.3: The estimated fund's average styles over the whole period covered.

To evaluate the fund's performance, we constructed a benchmark portfolio with similar style characteristics to that of the Halifax fund using the continuous change and dominant style approaches. Panel A of Table 5.21 below shows that, under a continuous changing style approach, Halifax fund underperformed its style mimicking benchmark by 0.16% per month (-1.92% per year). The statistic was even more alarming under the dominant style approach, where the underperformance is -0.24 per month (-2.88 per annum), statistically significant at the 10 % level. Furthermore, the variance of Halifax fund's return is higher than its benchmark, implying that the benchmark index is far more diversified than the Halifax fund. Hence, fund manager's selection skills have made the fund riskier compared to its style mimicking benchmark.

Next, we turn our attention to assess the extent to which the performance is attributed to investment styles and the active management effect. If we are assuming that the four style mimicking portfolios are the only source of variation in fund's return, then the R-square value would represent the performance attributed to the fund's style and (1- R-square) represents the performance attributed to the fund manager's selection skills (Sharpe, 1992). Thus, the higher (1- R-square) proportion indicates a relatively more active management. The R-square is calculated as $R^2 = 100 * (1 - \frac{variance\ of\ AR_t}{variance\ of\ R_t})$. Figure 5.4 provides a graphical summary of the results obtain under the continuous changing and dominant style. Under both approaches, around 82% of the monthly variation in funds' return can be explained by the fund exposure to the benchmark style indices, while around 18% is attributed to fund manager stock selection skills.

Based on these results, the Halifax fund underperformed its designated style benchmark. The fund manager's stock selections also made the fund riskier compared to its style benchmark. The inferior performance of the fund is also likely to deteriorate further when the cost of management fees is taken into account.

Table 5.21: Descriptive statistics of average monthly performance of Halifax UK Growth

	Panel A			***Panel B***		
	Continuous Changing			***Dominant***		
Series	***Average Returns***	***Std - Error***	***t-stat***	***Average Returns***	***Std - Error***	***t-stat***
$\boldsymbol{R_t}$	0.49%	4.10%	1.39	0. 49%	4.10%	1.39
$\boldsymbol{RB_t}$	0.66%	3.70%	2.02*	0.73%	3.74%	2.25*
$\boldsymbol{AR_t}$	-0.16%	1.76%	-1.05	-0.24%	1.7%	-1.61

Notes: The symbols used imply the following: R_t is the monthly realized return of Halifax UK growth. RB_t is the benchmark portfolio constructed in two ways: In the first (continuous changing style), the benchmark is the sum of the product of the yearly factor loadings and their corresponding investments style mimicking portfolio. In the second (Dominant style), the benchmark is the sum of the product of the average factor loadings and their corresponding investments style mimicking portfolio. AR_t is the monthly return difference between R_t and RB_t. * Indicates significance at the 5 percent level.

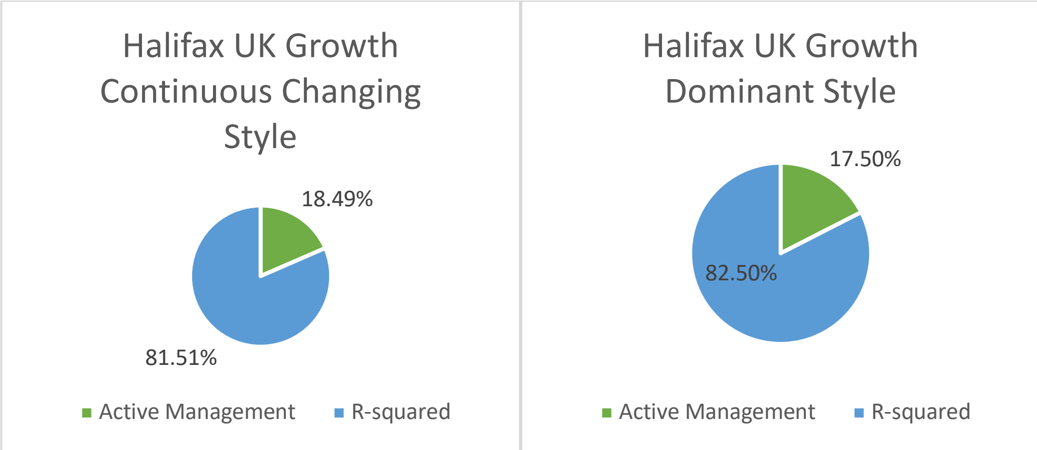


Figure 5.4: Halifax UK Growth performance attribution.

5.4.2. F&C Responsible Equity Growth

The F&C Responsible fund is classified as a UK flex-cap fund by Morningstar with assets under managements of £377 million as of December 2018 (Morningstar, 2018, b). Figure 5.5 illustrates the F&C fund's return exposure to style mimicking portfolios using a continuous changing style approach between 2005 and 2017. It clearly shows that before 2008, the biggest portion of funds were allocated to small-value stocks (almost 50%). However between the period 2008 to 2013, the fund steadily increased its exposure to big-oriented stocks. After 2013 the biggest style exposure chosen by the fund's manager is the big-value style. Indeed, our analysis shows that F&C fund's managers rotated its investment style across different market capitalization stocks over the course of the studied period.

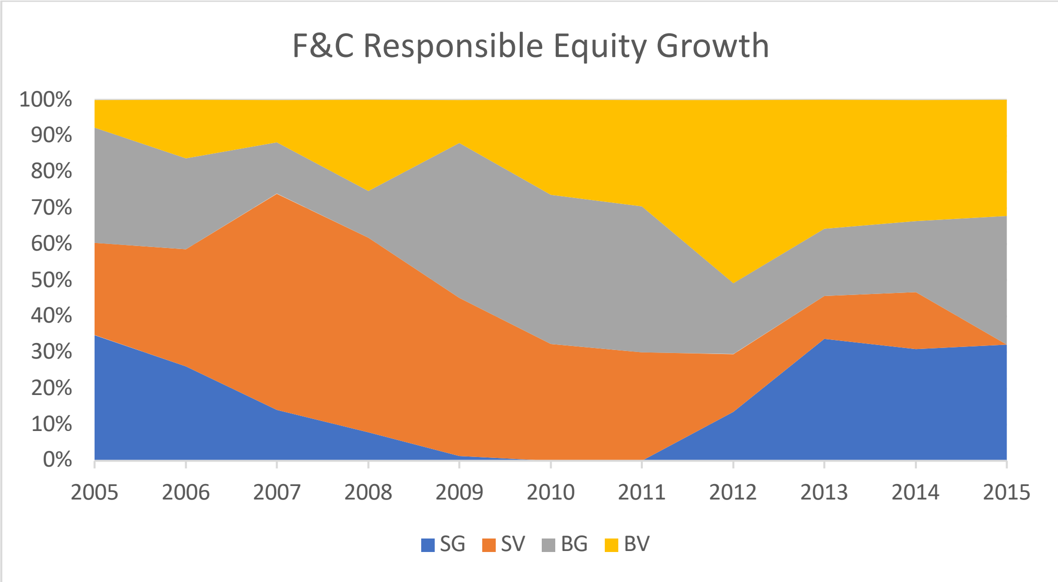


Figure 5.5: The estimated fund's changing styles over the period covered.

The bar chart in Figure 5.6 implies that the F&C Responsible fund has approximately equal exposure to selected style mimicking portfolios using the dominant style approach. Thus, to generate a return similar to the F&C Responsible fund, one would invest 18% in a small-growth benchmark index, 29% in a small-value benchmark index, 27% in a big-growth benchmark index, and 25% in a big-value benchmark index. The important conclusion from Figure 5.5 and 5.6 is that average changing style (Dominant) give us little information about the true fund's exposure when fund managers rotate their fund's exposure across different investment styles.

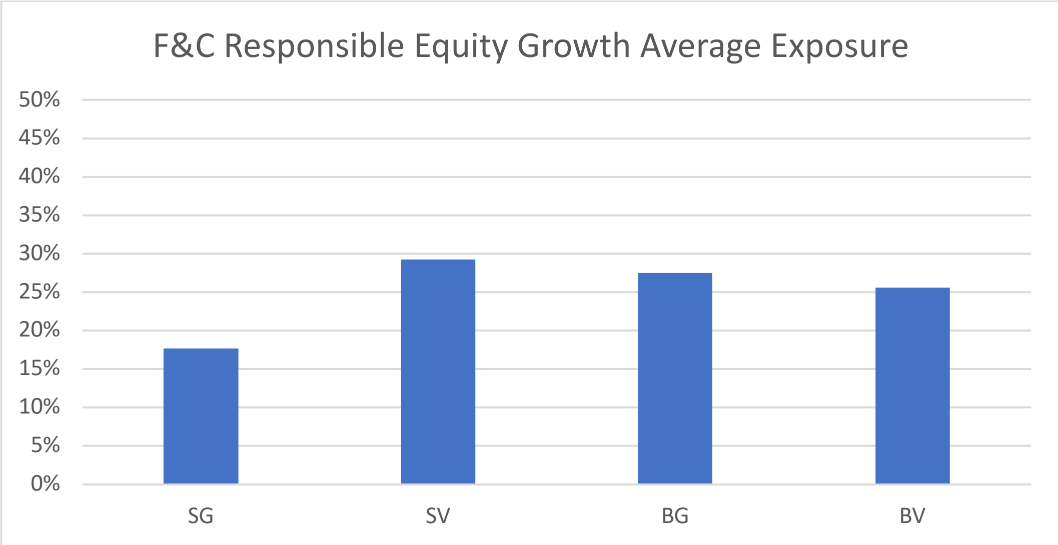


Figure 6: The estimated fund's average styles over the whole period covered.

Table 5.22 below shows that, under a continuous changing style approach, the monthly average return is identical for both the F&C fund and its designated benchmark. However, the average monthly return is statistically insignificant at any acceptable level. The standard deviation is marginally lower for the F&C fund, hence the manager's selection skill has enabled the F&C fund to achieve lower risk compared to its benchmark. In contrast, under a dominant style approach, the F&C fund has an average monthly return of 0.56%, while the benchmark with an equivalent style achieved a monthly return of 0.76% with a marginally lower standard deviation. The difference between the two returns, is however statistically insignificant at any conventional level. This result suggests that the performance is greatly influenced by whether a continuous changing style or dominant style approach is employed. Although a continuous changing style seems to be more appropriate to evaluate the fund's style performance, the dominant style approach might represent the experience of investor's more accurately.

Table 5.22: Descriptive statistics of average monthly performance of F&C Responsible Equity Growth.

	Continuous Changing			***Dominant***		
Series	***Average Returns***	***Std - Error***	***t-stat***	***Average Returns***	***Std - Error***	***t-stat***
$\boldsymbol{R_t}$	0.56%	4.08%	1.58	0. 56%	4.08%	1.58
$\boldsymbol{RB_t}$	0.56%	4.29%	1.5	0.76%	4.03%	2.17*
$\boldsymbol{AR_t}$	0.003%	1.83%	0.01	-0.20%	1.7%	-1.34

Notes: The symbols used imply the following: R_t is the monthly realized return of Halifax UK growth. RB_t is the benchmark portfolio constructed in two ways: In the first (continuous changing style), the benchmark is the sum of the product of the yearly factor loadings and their corresponding investments style mimicking portfolio. In the second (Dominant style), the benchmark is the sum of the product of the average factor loadings and their corresponding investments style mimicking portfolio. AR_t is the monthly return difference between R_t and RB_t.

Figure 5.7 shows that under the dominant style approach, 82.5% of the monthly variation in the F&C fund's return can be explained by a benchmark with the same style exposure. The remaining 17.5% is attributed to fund manager stock selection skill. Under a continuous changing approach, only 79.21% of the variation is explained by the benchmark index. The low R-Squared value might be due to the fact that the variance of the continuous changing style benchmark is more volatile than that of the dominant style benchmark. This result might be evidence of a style rotation strategy. In other words, stronger management activities can be traced in a continuous changing style. However, the F&C fund managers do not seem to add value through active management.

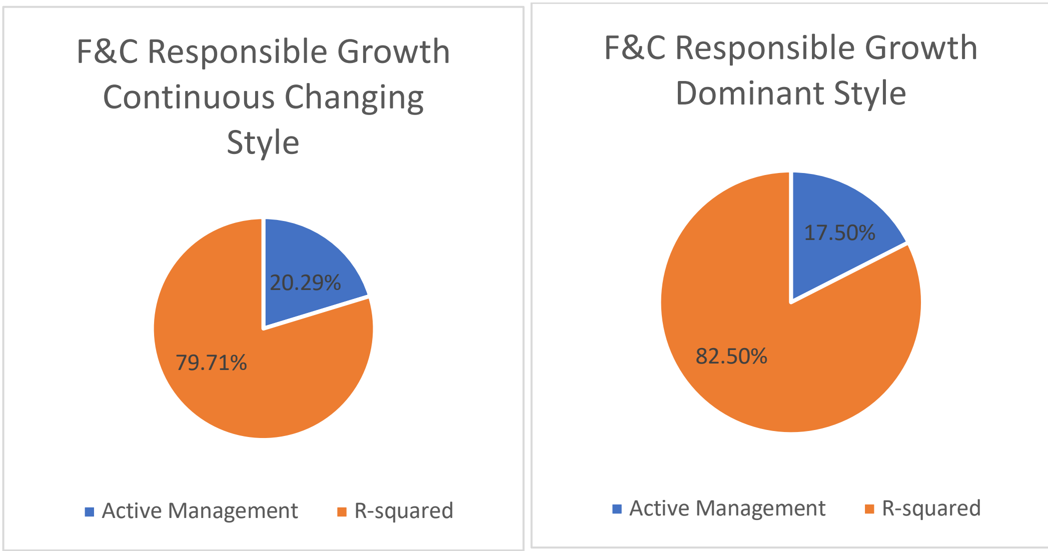


Figure 5.7: F&C Responsible Growth performance attribution.

5.5. Conclusion

This chapter has carried out a set of empirical tests on UK-equity fund's style-adjusted performance. The core objective has been to test Sharpe's asset-classes model and to highlight its ability in explaining UK-equity fund returns. We decomposed fund's return into size and growth-value dimensions and explored whether funds' performance differed across styles, scrutinising their ability to generate an abnormal return on style-based performance. The vast majority of prior studies showed that fund managers are unable to generate abnormal returns, and that they are more likely to underperform the passive benchmark. However, one would expect fund managers who effectively follow a successful investment strategy to earn enough returns to justify their management fees. Furthermore, fund's performance is very likely to be mis-evaluated when an inappropriate benchmark is used.

In this chapter, funds' returns are regressed against four investment style dimensions that have been documented in prior empirical research. The factor loadings are computed using 36 month rolling regressions. Although this window is sufficient to provide an accurate estimate of the factor loadings, it presumes that style is constant within the estimation period. Two procedures for style identification have been considered, the first is the continuous changing style, whereby UK-equity funds were grouped into four stylized portfolios based on the yearly variation of asset-classes factor loadings. In the second approach, we considered dominant style, whereby we grouped funds together based on their average changing style across the whole sample period. Under both approaches, we found that funds have a tendency to favour big-oriented stocks. Thus, the bulk of funds do not deviate from the market index (FTSE100). One possible explanation for this is that fund managers are aware of the difficulty in achieving long-term abnormal performance in an efficient market, hence they simply track the market index despite claiming otherwise. Thus, fund managers seem to involve themselves in window dressing activity to improve ex-post performance, with big stocks more likely to be included in their portfolio as they are easier to justify to investors (Lakonishok et al. 1991).

With regards to performance more generally, the results reveal that on average UK-equity funds neither underperformed nor overperformed their designated style benchmark. This finding is consistent with previous studies, such as Chan et al. (2002), Dimson et al. (2003) and Cuthbertson (2008) who report similar findings after controlling for size and growth-value factors' exposure. These results put doubt on fund managers' ability to earn an abnormal performance from their investment strategies. Hence, on average, the UK equity funds' managers have no added value attributed to fund managers' stock selection skill. However, similar to the finding of Brookfield (2013), we documented some evidence whereby investors can enhance their risk/style adjusted performance by investing in funds tilted toward small-value stocks, and avoiding funds with a growth-oriented investment style. Furthermore, our results suggest that UK-equity funds have generated lower dispersion than their benchmarks and the market index (FTSE 100). Thus, fund managers' selection skills have made the funds less risky compared to their benchmarks. However, Ben Dor and Jagannathan (2003) argued that grouping funds in portfolio significantly reduces their variance. Hence, we cannot conclude that fund managers, on average, generated lower dispersion than the relevant style benchmark.

The second part of the chapter discuss whether the UK-equity ethical funds' returns can be explained by fund investment style. Our analysis showed that ethical funds' exposure to small-oriented stocks has declined steadily since the 2007 financial crisis. This result largely contradicts previous research on ethical funds' holdings (see for example Hamilton et al., 1993; Gregory et al., 1997; and Goldreyer et al., 1999). In relation to ethical fund performance, under a continuous changing style approach, growth-oriented ethical funds have earned lower return than their style benchmark indices. Furthermore, when performance is compared with conventional funds, it can be seen that ethical fund managers did worse than conventional fund managers on a style-adjusted basis. This result is consistent with previous empirical work on ethical funds in the UK market (Luther et al., 1992; Mallin et al., 1994; and Gregory et al., 1997). However, our analysis showed that the disappointing performance cannot be blamed on ethical funds exposure to 'small firms' risk. With regards to a dominant style approach, the results tentatively showed no support for ethical funds underperformance compared to its conventional peers. Thus, we can conclude that the performance is influenced by whether a dominant or a continuous changing style is used in the portfolio formation method.

In conclusion, on average, both ethical and conventional funds exhibit no significant abnormal performance on a style adjusted basis. Furthermore, the style/risk adjusted performance is likely to be below the average 1.95% per year fees charged by said funds (see chapter 3 for mor information). Thus, active managements cannot justify the higher fees that they charge relative to cheaper passive options. Finally, ethical investors are expected to pay a heavier financial price for being ethical than their conventional counterparts.

Chapter 6

The Performance of Ethical and Non-Ethical Funds: Event Time Results

6.0. Introduction

In the previous chapter, we analysed UK equity funds' performance using a return-based style analysis approach. The abnormal returns are measured as the returns in excess of a style-benchmark that quantifies the performance of the funds more effectively than a generic market index. The main findings were that UK equity fund managers generally underperform their respective style benchmark and that UK equity fund managers do not, on average add any value above that attributed to a fund's investment style. However, Liu and Strong (2008) pointed out that evaluating the performance of an investment style based on formulating a single-period portfolio return over a multi-period holding horizon is misleading and does not capture the true return from a buy-and-hold strategy over the investment holding period. Thus, portfolio rebalancing inherent in measuring investment performance might introduce false inference and does not correspond to the returns that investors have actually accumulated by the end of the holding period investment strategy. In this chapter, we continue exploring the return performance of UK-equity funds by closely examining funds' investment style, using the event studies framework and Liu and Strong's (2008) method for calculating the profits of a short-long run strategy. Thus, within this context, this chapter aims to address the following empirical question. If fund managers produce alpha attributed to fund's investment style, can investors exploit an ex-ante investment style strategy and how should they frame their investment horizon between competing strategies?

6.1. BHAR Results for Conventional (Non-Ethical) Funds

Panel A of Table 6.1 reports the equally weighted mean BHARs of funds with the highest exposure to small-growth stocks against a small-growth characteristic-based reference portfolio across different holding periods. The mean BHAR declined from -4% after 12 months to -9% after 24 months, significant at the 1 per cent level. The rate of decline stabilised around the level of -11% between 24 and 48 months before falling to its lowest level of -28% after 60 months, and all are significant at the 1% level. The median performance is slightly worse than the mean performance except for the 12- and 48-month investment horizons. Meanwhile, the mean BHAR distribution is symmetrical, with a skewness value close to zero throughout the investment horizons (shown in column 6). Nonetheless, the BHAR return exhibits high leptokurtic properties, thus the BHAR distribution implies fat tails relative to the normal distribution and the extreme values of returns are likely to influence the power of the test statistic. Although the wild bootstrap test statistic is effective in handling outlier, we report the mean return after winsorizing the abnormal performance at the 1% and 99% level. The reported truncated means show trivial differences in comparison to the overall BHAR mean. Similar statistical significance is found in both the skewness-adjusted and kurtosis preserved wild bootstrap. The skewness adjusted t-statistics show no significant difference from the conventional t-test, and the p-values using the kurtosis preserving wild bootstrap (pv1) are consistent with the conventional t-test at all investment horizons. This result suggests that if investors had chosen to systematically invest in small growth funds, they would have generated negative returns ranging from 5 % to 28% depending on their investment horizon. Yet, their performance is most likely to deteriorate further with the introduction of transaction costs and management fees. Thus, one would argue that, on average, small-growth fund managers are incompetent in tracking the performance of their specific style benchmark and possess poor stock selection skills.

Panel B of Table 6.1 presents the mean BHARs results from matching against the market portfolio (FTSE 100). This allows us to attribute performance to investment style, with the mean BHARs representing the active management effect, or funds' average selection return. The results show that the small-growth UK equity funds' managers have failed to beat the market benchmark and generate abnormal performance. At all investment horizons, the mean BHARs are statistically insignificant at any conventional levels, suggesting that there is no difference in return between average small growth funds and the FTSE 100. Both medians and the truncated means have a similar pattern to the mean performance, indicating that the variation in the average abnormal return is symmetrical. Although the distribution exhibits high leptokurtic properties, the corrected p-values using 1000 random samples show that the BHAR's statistical properties are well specified, and that a high degree of kurtosis has no direct effect on the power specification of BHAR. This finding suggests that small-growth funds' managers are unable to capture the small size premium that has been documented in common stock returns, and therefore that small growth funds' managers cannot justify the higher fees that they charge relative to cheaper passive options. It is worthwhile to note that the bulk of our sample funds started in 2005 and that their style exposures were re-estimated in 2010. These two investment horizons cover significant economic fluctuations; the global financial crisis of 2007 and the European sovereign debt crisis of 2011, economic fluctuations that one might expect to contribute to small size stocks performing particularly badly, with liquidity drying up and small cap companies therefore struggling to raise capital.

Panel A of Table 6.2 presents the results from the use of an equally weighted mean BHARs of funds with highest exposure to small-value stocks against a small-value characteristic-based reference portfolio. The mean BHAR starts off as being significantly negative at the 1% level for the one- and two-year holding periods, recording -4% and -7% respectively. By the third year, the mean BHAR becomes a positive 5%, statistically significant at the 5 % confidence level. It then increases to 13% after 4 and 5-year periods (statistically significant at the 1% confidence level). The results also indicate a wide variation of average abnormal returns amongst funds given the significant differences between the means and the medians, in particular for the 4 and 5-year periods, which suggest that there is more variability on the left side of the distribution, specifically below the 1 % level. Columns 6 and 7 show that the returns distribution is symmetrical but highly leptokurtic for the 2-year period. The test statistics of the skewness-adjusted and kurtosis preserved wild bootstrap is consistent with the conventional t-test.

Accordingly, we can conclude that on average the long run performance of a small-value portfolio is reliably positive for a holding period of 3 to 5 years. Thus, on average a zero initial investment which would be achieved by taking a long position in the small-value funds and a short position in the reference portfolio, would have resulted in a gross profit for investors of 5% by the end of 3 years and 13% by the end of a 4- and 5-year holding period.

Table 6.1: Small-Growth Buy and Hold Abnormal Return (BHAR) event time returns.

Panel A. using small-growth characteristics-based reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adjt-statistic*	*q-1%*	*q-99%*	*pv1*
12	-0.05	-4.01	-0.04	0.09	-1.44	4.49	-0.04	-5.08	-2.40	2.38	0.00
24	-0.09	-4.35	-0.10	0.16	-0.92	3.12	-0.09	-5.15	-2.31	2.14	0.00
36	-0.11	-4.35	-0.13	0.18	-0.27	1.37	-0.10	-4.58	-2.28	2.14	0.00
48	-0.12	-4.55	-0.07	0.19	-0.44	-0.47	-0.11	-4.96	-2.25	2.43	0.00
60	-0.28	-8.06	-0.29	0.26	0.15	2.31	-0.28	-7.63	-2.26	2.41	0.00

Panel B. using FTSE100 reference benchmark.

Holding period	**Mean**	**t-statistic**	**Median**	**SD**	**Skewness**	**Excess Kurtosis**	**Truncated Mean**	**SK-Adjt-statistic**	**q-1%**	**q-99%**	**pv1**
12	0.01	1.17	0.00	0.08	-0.67	3.07	0.01	1.11	-2.34	2.39	0.26
24	0.02	1.06	0.02	0.16	-0.30	2.03	0.02	1.04	-2.26	2.16	0.30
36	-0.02	-0.58	-0.05	0.24	0.06	1.55	-0.02	-0.58	-2.52	2.13	0.58
48	-0.04	-1.11	-0.10	0.25	0.61	1.55	-0.05	-1.06	-2.85	2.38	0.30
60	0.02	0.54	0.00	0.33	0.87	3.78	0.00	0.57	-2.31	2.23	0.60

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness, kurtosis, and a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 is the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

One possible explanation for the long-run abnormal performance is that small-value funds' managers possess superior skills that allowed them to beat their style specific benchmark. In particular, fund managers were able to identify loser stocks in small value's style during the economic downturns, which then generated long-run abnormal returns when the prices of the loser stocks reverted to fair values.

Panel B of Table 6.2 provides a summary of the results from a comparison against market portfolio. It appears that a small-value oriented portfolio exhibits a positive abnormal performance relative to the FTSE 100 at all investment horizons. The magnitude of abnormal performance ranges between 2% to 4% for the first 4 years then surges to 13% after a 60-month holding period. However, the t-statistics associated with the mean BHAR is statistically insignificant except for the 60-month holding period, which is highly significant at the 1% level. The BHAR distribution is skewed to the right, and highly leptokurtic, specifically for the 4 and 5-year periods. This is typically because the sample BHAR includes outliers from the right tail of the distribution, and these outliers boost the standard deviation and lower the t-statistic making it harder to reject the null hypothesis. The skewness-adjusted t-statistics and the kurtosis preserved wild bootstrap show similar significance level to that observed from the conventional t-test. Whereby, the abnormal return remains highly significant for the 60-month period. In effect, we can conclude that on average the long run performance of a small-value portfolio is positive for a holding period beyond 60 months. However, the generation of positive long-run abnormal return does not necessarily imply that investors are better off investing in small value funds. This is because small value funds might contain residual risk relative to the market benchmark, which may or may not be correlated with the factor returns.

Panel A of Table 6.3 reports the result of the mean BHAR, which is derived from big-growth funds against big-growth characteristic-based reference portfolio. The results suggest that the big-growth funds consistently underperform their benchmark at all investment horizons. The performance declines gradually as we increase the holding period, with a steady decline of approximately 5% per year. The median and truncated mean are more negative than the overall BHAR mean for a holding period of more than 3 years. Besides, the BHAR distribution exhibits high kurtosis at all investment horizons except for the 2-year holding period. Furthermore, the size of the conventional t-test is troublingly high across all investment horizons. Although the skewness-adjusted test statistic reports a lower value, specifically for a long-run investment horizon, the size of the t-statistic remains very high. Similarly, the kurtosis preserving wild bootstrap shows that the underperformance is highly significant at all investment horizons. One possible explanation is the cross-sectional dependence which was observed by Mitchell and Stafford (2000). According to these authors, the BHAR assumes equal variance and pairwise covariance across all sample funds' abnormal returns. However, Bernard (1987) reports that the pairwise cross-sectional correlation increases with both sample size and return horizon. Since most funds in our sample are tilted toward a big-growth investment style[12] and our investment horizon is up to 5 years,

---

[12] See Chapter 5 for more details.

Table 6.2: Small-Value Buy and Hold Abnormal Return (BHAR) event time returns.

Panel A: using small-value characteristics-based reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	-0.04	-3.21	-0.04	0.08	-0.08	0.27	-0.03	-3.25	-2.32	2.37	0.00
24	-0.07	-3.63	-0.08	0.14	0.43	2.89	-0.08	-3.35	-2.46	2.38	0.00
36	0.05	1.8	0.05	0.19	-0.3	-0.2	0.05	1.75	-2.41	2.46	0.05
48	0.13	2.78	0.23	0.33	-0.59	-0.23	0.15	2.56	-2.18	2.3	0.01
60	0.13	2.77	0.22	0.34	-0.94	0.47	0.17	2.41	-2.28	2.54	0.01

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	0.02	1.71	0.01	0.10	0.57	1.82	0.02	1.80	-2.44	2.21	0.10
24	0.04	1.41	0.02	0.20	0.83	1.20	0.03	1.50	-2.31	2.15	0.16
36	0.01	0.49	-0.02	0.20	0.84	1.49	0.00	0.52	-2.45	2.41	0.63
48	0.02	0.71	-0.06	0.25	1.22	1.77	0.00	0.76	-2.71	2.27	0.51
60	0.13	2.96	0.06	0.32	1.83	5.10	0.09	3.75	-2.43	2.42	0.00

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean and truncated t-test based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 is the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

even a small degree of cross-correlation in our sample can inflate the t-statistics significantly. Thus, the wild bootstrap procedure fails to account for cross correlation. Nonetheless, the BHAR results suggest that investors with a preference for big-growth investment funds have achieved lower returns relative to the benchmark index, and the underperformance worsened the longer their investment horizon is. Consequently, big-growth funds' managers do not possess sufficient skill to cover costs, but rather inferior skills that reduce their fund's returns.

Panel B of Table 6.3 presents the results from matching against the FTSE100 index. Unlike the result reported in panel A, the mean BHAR is small in absolute magnitude and varies from 0% to 8%. Specifically, big-growth fund managers have, on average, mimicked the performance of the FTSE100 for an investment horizon of up to 3 years. Thereafter, they were able to produce abnormal returns relative to the market benchmark (FTSE100) by 2% and 8% for the 4 and 5-year holding periods, respectively. However, the conventional t-test shows that the abnormal return is statistically insignificant except for the 5-year holding period. The mean BHAR is skewed to the right and is highly leptokurtic at all investment horizons. However, the adjusted-test statistic and the kurtosis preserving wild bootstrap show similar pattern to the conventional t-test, being highly significant only at the 60-month holding period.

Generally, the performance of the big-growth mutual funds looks better when returns are measured relative to the market index. However, this statement is not absolutely true since employing the characteristic based reference portfolio (Panel A) allows for cross-sectional variation in expected return. This can be seen when comparing the descriptive statistics in Panel A and B of Table 3. The variances obtained from a big-growth reference portfolio are, for example, smaller than those obtained from the market benchmark, particularly beyond a 36-month horizon.

In conclusion, big-growth funds' managers deliver market benchmark-adjusted returns in the short- and medium-term investment horizon. In the long-run, investors would experience an abnormal return of 8% for a holding period of 5 years. The BHAR distribution suggests that there are a few funds managers who produce abnormal returns. However, their performance is crowded out by the majority of managers with a performance level below the market benchmark-adjusted returns.

Panel A of Table 6.4 provides a summary of the BHARs distribution for big-value tilted funds over a 12 to 60-month holding period. In the short-run, the result suggests that big -value funds delivered negative returns of -2% and -9%, significant at the 1% level for the one- and two-year horizons, respectively. Thereafter, there seems to be some recovery as the mean BHAR turns positive, achieving 10% and 15% over the reference portfolio for the 4 and 5-year holding period, statistically significant at a 1% level. Both the median and the truncated mean are substantially below the BHAR mean at all investment horizons, and it might suggest that the BHAR results might be driven by the outliers. The application of skewness-adjusted t-statistic and kurtosis preserving wild bootstraps show that the conventional t-test is biased toward negative performance. However, the rejection rates remain the same except for the 12-month period, where we rejected the null hypothesis at the 5% level instead of the 1% level reported in the student t-test. In conclusion, on average, big-value funds exhibit significant positive long-run abnormal performance on a style adjusted basis. Thus, some fund managers who effectively follow big-value investment strategies are likely to earn economic rent that justifies their management fees.

Panel B of Table 6.4 shows similar result of the big-value mean BHARs using the market benchmark. The abnormal performance is statistically significant at the 1% level beyond a 36-month horizon, where the size of the abnormal performance varies between 9% and 12% for a 48 and 60-month holding period, respectively. The BHAR distribution is positively skewed and highly leptokurtic at the first- and second-year holding period.

Neither the skewness-adjusted bootstrap nor the kurtosis preserving bootstrap indicate that the conventional t-test findings are being driven by outliers. The one exception is that the t-test of the skewness-adjusted wild bootstrap which reports significant positive performance for the 36-month horizon at 5% level, while the t-test of the kurtosis preserving wild bootstrap and the standard t-test report otherwise. However, it is worth noting that, the BHAR distribution is symmetric for the 36-month holding period. Overall, the results suggest that big-value oriented funds' managers can generate abnormal performance relative to the FTSE100 market index. Beyond a 36-month investment horizon, investors are expected to gain a gross abnormal return of up to 12%.

Table 6.3: Big-Growth Buy and Hold Abnormal Return (BHAR) event time returns.

Panel A: using big-growth characteristics-based reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	-0.05	-11.18	-0.06	0.06	-0.02	2.32	-0.05	-11.25	-2.48	2.34	0.00
24	-0.11	-14.76	-0.11	0.10	0.19	0.20	-0.11	-13.77	-2.22	2.59	0.00
36	-0.17	-17.71	-0.17	0.13	0.81	3.13	-0.17	-11.66	-2.39	2.41	0.00
48	-0.23	-18.64	-0.25	0.17	1.40	3.21	-0.25	-7.15	-2.23	2.55	0.00
60	-0.28	-20.34	-0.30	0.19	0.90	2.16	-0.29	-11.55	-2.15	2.41	0.00

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	0.00	-0.68	0.00	0.07	0.30	3.12	-0.01	-0.68	-2.28	2.23	0.49
24	0.00	0.42	-0.01	0.09	0.70	3.31	0.00	0.43	-2.35	2.40	0.69
36	0.00	-0.18	-0.02	0.14	1.37	4.76	-0.01	-0.16	-2.44	2.31	0.86
48	0.02	1.47	-0.02	0.20	1.54	3.80	0.00	1.57	-2.40	2.70	0.14
60	0.08	4.74	0.04	0.22	1.16	2.68	0.05	5.37	-2.25	2.16	0.00

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean and truncated t-test based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 is the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

Table 6.4: Big-Value Buy and Hold Abnormal Return (BHAR) event time returns.

Panel A: using big-value characteristics-based reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	-0.02	-2.12	-0.04	0.09	1.86	4.98	-0.04	-1.75	-2.39	2.25	0.04
24	-0.09	-3.84	-0.15	0.19	1.15	1.02	-0.11	-3.14	-2.40	2.46	0.00
36	0.00	-0.18	-0.06	0.21	1.14	1.73	-0.03	-0.15	-2.58	2.36	0.86
48	0.10	2.64	0.03	0.31	1.21	1.84	0.06	3.00	-2.31	2.41	0.01
60	0.15	4.27	0.10	0.30	0.64	0.20	0.13	4.75	-2.23	2.36	0.00

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
12	-0.01	-1.30	-0.02	0.07	1.25	2.13	-0.02	-1.19	-2.27	2.29	0.21
24	0.01	0.95	-0.02	0.11	1.28	1.86	0.00	1.02	-2.58	2.23	0.36
36	0.03	1.59	0.01	0.15	0.67	0.34	0.02	1.67	-2.28	2.52	0.12
48	0.09	3.03	0.05	0.24	0.83	0.91	0.07	3.35	-2.41	2.70	0.00
60	0.12	4.20	0.08	0.23	0.38	-0.32	0.11	4.48	-2.32	2.25	0.00

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

6.1.1. Summary of Conventional (Non-Ethical) Funds BHAR Results.

To summarize our result so far: We conducted BHARs methodology to measure and explain the long-run performance of UK equity funds after controlling for the funds' size and value-growth characteristics. We explored whether funds' performance differs across styles and scrutinised their ability to generate abnormal returns for an investment horizon of 1, 2, 3, 4, and 5 years.

The results using the characteristic based reference portfolio are different from those using the market benchmark (FTSE100). Clearly the relative long-run performance is benchmark dependent. However, one would expect the characteristics-based reference portfolio to capture systematic risk beyond the market benchmark. Several studies in the UK have documented the effects of investment style on fund performance. For example, Quigley and Sinquefield (2000), and Brookfield et al. (2013) both favour value-oriented over growth-oriented funds on a long-term risk-adjusted basis.

Although the long-term abnormal returns reported in our study conform closely with the existing literature, the short-term abnormal returns do not. Our empirical findings suggest that relative to the reference portfolio, value-oriented funds underperform, or at least disappoint investors in the short-run. Meanwhile, value-oriented funds' managers deliver positive abnormal returns for an investment horizon beyond 36 months. However, these performance gains are most likely to fade away with the introduction of management fees.

In contrast, the underperformance is more pronounced for growth-oriented funds and the longer the holding period, the more severe growth-oriented investment can become. Our evidence further shows that underperformance is not driven by a size effect, with small-growth unit trusts performing as badly as big-growth funds. Because ignoring characteristic-based benchmark introduces biases in the performance measurement, as Pettengill et al. (2013) suggest, we also report performance relative to passive portfolio benchmark (FTSE100).

The general pattern is that there are no significant differences in returns on the short-and medium-term horizon. However, positive abnormal performance is detected for value-oriented and big growth funds, but only at the 60-month investment horizon. On the other hand, when we compared the statistical properties of the BHAR's distribution using a characteristic-based reference portfolio and the market index (FTSE 100). The results show that the standard deviation of a value-oriented mean relative to the market index is somewhat lower than the standard deviation of the value-oriented mean relative to a characteristic-based reference portfolio, specifically at investment horizons beyond 36 months. This questions the validity of the metric used to capture expected return, and one may argue that using the market benchmark would be more appropriate. However, the distribution of returns tends to be more normally distributed using the characteristic-based reference portfolio compared with the FTSE 100.

In conclusion, our results suggest that, on average, value-oriented fund managers possess sufficient stock selection skills that allow them to deliver positive abnormal performance in the long-run. In an attempt to explain such abnormal performance Conrad, et al. (2003) argue that the variation between the fund characteristics and fund returns can be explained by data snooping biases. Moreover, Ali et al. (2003) show that the value effect is driven by stocks with higher idiosyncratic risk, higher transaction costs, and lower investor sophistication. Most importantly, models of expected return such as the Fama and French 4 factor model may be a more appropriate measure than the characteristic based reference portfolio.

Furthermore, our hypothesised event (investment style) is likely to be a non-random occurrence. For example, fund managers might involve themselves in window dressing and style rotation activities to improve ex-post performance. And since our analysis presumes that a fund's style is constant within the test period, there is a danger that the small-value funds and their characteristic-matched reference portfolio differ systematically in their expected returns. Nonetheless, Pettengill, et al. (2013) pointed out that investors are only concerned about realized return and realized risk rather than expected, and investors evaluate the success of their fund managers based on a self-reported benchmark stated in the fund's prospectus rather than risk-adjusted returns. Taken together, these considerations suggest that the long-run abnormal performance of value-oriented funds could be conclusive.

On the other side of the spectrum, on average, growth-oriented fund managers delivered significant underperformance at all investment horizons. One possible explanation is that the presence of industry clustering and calendar time clustering among growth-oriented funds managers. Particularly, in economic downturns, fund's managers seem to be involved in buying overvalued growth stocks as they are easier to justify to investors. If that is the case, then at certain times, many growth funds buy stocks which are irrationally high in price, since these prices are unsustainable, then negative abnormal returns will be realized in the long-term when prices of overvalued stocks revert to fair values. In effect, individual funds' abnormal returns are cross sectionally correlated in the BHAR calculation, and this explains the troubling size of the t-statistics of all the mean BHARs.

6.2. CARs Results for Non-Ethical Funds.

Panel A in Table 6.5 to Table 6.8 present the results of the CAR derived from equally weighted stylized portfolios against equivalent characteristics-based reference indices. The statistical behaviour of unit trusts' returns for every 12-month period for up to 60 months reveals that the CARs returns are smaller than the BHARs in terms of absolute magnitude. Although the BHAR method systematically magnified performance at all investment horizons, the divergence in CARs returns relative to BHARs becomes increasingly significant at horizons beyond 36 months. Figure 6.1 confirm these findings.

These results conform closely with earlier studies, such as Fama (1998), Mitchell and Stafford (2000) and Gompers and Learners (2003) who argue that the BHAR method tend to inflate under/overperformance, even if it only occurs in a particular period. In effect, one may conclude that the BHAR method suffers from the compounding problem. However, by closely comparing the CAR and BHAR returns, specifically for value-oriented funds at the 36-month horizon, we notice that even when negative returns have been realized at shorter horizons, BHAR returns bounced back at a faster and larger magnitude than CARs. This finding contradicts the results observed by Mitchell and Stafford (1997), where the compounding effect in BHAR was found to disguise the actual speed of performance adjustment.

When comparing the test statistics (standard t-test) obtained using the CAR and BHAR methods, we observed a similar statistical significance of the estimated performance using the two approaches. Thus, the performance observed from BHAR remain detectable by CAR at all investment horizons. This result is inconsistent with Barber and Lyon (1997) who argue that CAR will tend to positively bias test statistics, while BHAR produces negatively biased test statistics. They suggest several reasons for this bias, including rebalancing bias, new listing bias and skewness bias. However, as we discussed earlier in chapter 4, our reference portfolio is constructed in a way that eliminates or reduces such biases. The returns of the characteristic-based reference portfolios are likely to be just as skewed as sample funds' returns. However, we observed some degree of skewness in the CARs distribution, specifically at short term horizons (12 and 24 months), except for small-growth funds where skewness is also observed at longer horizon (48 and 60 months). In contrast, as it can be seen in Figure 6.2, the BHAR distribution exhibited a lower degree of skewness, except for the big-value funds, where the distribution is positively skewed but not at the 60-month horizon. This finding is by and large inconsistent with Fama (1998), who advocated shorter intervals in event studies, since the distribution of returns tends to be more normally distributed in the short run compared with the long run.

The general pattern of the fourth moment shows that the CAR distribution is highly leptokurtic, especially for the 12- and 24-month holding periods. Except for small-growth funds, the distribution is severely fat-tailed across all investment horizons. Thus, the lower means in small-growth funds appears to be due to a few fund managers with extremely low returns. While positive skewness with fat-tailed distribution (i.e., big-value funds) indicates that there are a few funds manages who produce abnormal returns, their performance is hidden by the majority of managers whose performance is below the mean. Figure 6.3 shows that the distributions of BHARs exhibit fatter tailed returns than CAR returns. Similar results were reported by Kothari and Warner (1997), who found BHARs distribution to be asymmetric with a high value of kurtosis.

To scrutinize further whether the distributional properties of CAR and BHAR cause misspecified test statistics, we also reported the skewness-adjusted and kurtosis preserved wild bootstrap. Both BHAR and CAR produce well-specified test statistics in two-tailed test. The skewness-adjusted and kurtosis preserved wild bootstrap indicate that the rejection rates of the null hypothesis are consistence with the conventional t-test at all investment horizons. Thus, our nonparametric test shows little improvement in the specification, contrary to the results of Lyon et al. (1999).

Panel B of Table 6.5 to Table 6.8 present the result of CAR form a comparison against market portfolio (FTSE100). It shows that much of the performance identified using characteristics-based reference portfolios disappears when the market index is used as a benchmark. Similar to the BHAR results but of smaller magnitude, the abnormal performance is identified for the value-oriented and big growth funds, but only at the 60-month investment horizon, while there are no significant differences in returns on the short-and medium-term horizons. Most importantly, we observe significant improvement in the distributional properties of CAR. The CAR distribution is fairly symmetrical at all investment horizons and across the four investment strategies. However, high kurtosis is still present (though it is lower than the characteristic-based reference portfolio). Furthermore, the variance of difference between returns of the value-oriented funds and the market index is lower than that of the difference between the returns of the value-oriented funds and the characteristics-based reference portfolios. In particular, at a longer term horizon (48 and 60 months), the CAR standard deviations are substantially lower than those observed using the characteristics-based reference portfolios. This finding explains the lower power of the test in the market index approach, and might suggest that value-oriented funds manager simply track the market index despite claiming otherwise. When it comes to test specification, both the skewness-adjusted and kurtosis preserved wild bootstrap yield similar rejection rates to that observed in the conventional t-test at all investment horizons. Thus, we can conclude that the CARs provide a well-specified test statistic.

However, at the long-term horizon, both CAR and BHAR are prone to incremental misspecification in the standard deviation of the t-test. As we discussed earlier, this misspecification stems from the cross-sectional dependency of abnormal returns which was observed by Mitchell and Stafford (2000). Both CAR and BHAR assume that abnormal returns are independent and normally distributed. However, economy-wide, industry clustering and calendar time clustering would generate co-movements in fund returns that violate the assumption of independence of abnormal returns. For example, when the economy is in bad condition, a high percentage of funds may choose to involve in style rotation activities. This creates the issue of the clustering of abnormal returns, and violates the independence assumption. In other words, if the variance increase is purely fund specific, it would not affect the mean long-term abnormal returns. However, if the volatility is cross correlated in the sample funds, inferences would remain as treacherous as ever.

In conclusion, the choice between CAR and BHAR has been particularly difficult for two reasons. First, it is unclear which expected return benchmark is correct in estimating long-term abnormal return. Thus, a small error in the benchmark-adjusted returns can create significantly large differences. Second, although we believe that our non-parametric tests were successful in addressing the normality assumption, neither the skewness-adjusted nor the kurtosis preserved wild bootstrap were able to take account of cross-correlation of returns within the sample. However, CAR and BHAR measure different things. For example, 5-year CAR tests the hypothesis that the mean monthly abnormal return is zero during the 5 years, while the 5-year BHAR tests the hypothesis that the mean 5 years abnormal return is zero. Thus, in practice the BHAR method seems to be more representative of investor experience and is, in effect, our preferred method.

Table 6.5: Small-Growth Cumulative Abnormal Returns.

Panel A: using small-growth characteristics-based reference portfolio.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.04	-3.91	-0.04	0.08	-1.22	3.56	-0.05	-4.76	0.00	-2.36	2.50	0.00
24	-0.08	-4.00	-0.08	0.14	-0.93	2.91	-0.08	-4.68	0.00	-2.19	2.32	0.00
36	-0.09	-4.48	-0.10	0.15	-0.66	2.24	-0.10	-5.09	0.00	-2.40	2.65	0.00
48	-0.12	-4.48	-0.08	0.20	-1.28	2.95	-0.13	-5.66	0.00	-2.13	2.27	0.00
60	-0.22	-7.21	-0.20	0.23	-1.52	7.77	-0.23	-10.76	0.00	-2.36	2.13	0.00

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	0.01	1.45	0.01	0.08	0.25	3.91	0.01	1.48	0.18	-2.32	2.61	0.17
24	0.02	0.81	0.02	0.15	-0.21	3.01	0.01	0.80	0.44	-2.27	2.23	0.42
36	-0.02	-0.55	-0.03	0.21	-0.32	1.99	-0.02	-0.56	0.61	-2.32	2.52	0.61
48	-0.05	-1.67	-0.08	0.24	-0.76	3.00	-0.05	-1.78	0.12	-2.44	2.61	0.12
60	0.00	-0.13	0.01	0.27	-0.89	6.06	0.00	-0.15	0.91	-2.45	2.19	0.91

The columns show the CARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 is the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

Table 6.6: Small-Value Cumulative Abnormal Returns.

Panel A: using small-value characteristics-based reference portfolio.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.03	-2.86	-0.04	0.07	0.43	1.44	-0.03	-2.69	0.01	-2.26	2.33	0.01
24	-0.04	-2.66	-0.08	0.12	1.18	1.69	-0.04	-2.24	0.04	-2.45	2.38	0.01
36	0.04	1.70	0.03	0.15	-0.04	-0.04	0.05	1.70	0.09	-2.44	2.09	0.08
48	0.16	3.92	0.17	0.29	-0.19	-0.76	0.16	3.78	0.00	-2.13	2.39	0.00
60	0.08	2.75	0.12	0.22	-0.66	-0.03	0.11	2.50	0.01	-2.33	2.26	0.01

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	0.02	1.47	0.01	0.08	0.07	1.35	0.01	1.48	0.17	-2.33	2.67	0.16
24	0.01	0.67	0.01	0.16	0.07	0.73	0.01	0.67	0.48	-2.38	2.28	0.48
36	0.00	-0.15	-0.01	0.15	-0.01	0.48	-0.01	-0.15	0.89	-2.58	2.39	0.89
48	0.00	0.08	-0.05	0.18	0.46	-0.40	0.01	0.09	0.93	-2.47	2.13	0.93
60	0.07	2.87	0.06	0.18	0.43	0.59	0.08	3.05	0.00	-2.20	2.31	0.00

The columns show the CARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

Table 6.7: Big –Growth Cumulative Abnormal Returns.

Panel A: using big-growth characteristics-based reference portfolio.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.04	-9.49	-0.05	0.06	-1.20	5.17	-0.04	-12.07	0.00	-2.41	2.54	0.00
24	-0.08	-13.60	-0.08	0.09	-0.26	0.68	-0.08	-14.72	0.00	-2.38	2.48	0.00
36	-0.11	-16.10	-0.12	0.10	0.51	2.61	-0.11	-12.95	0.00	-2.32	2.28	0.00
48	-0.16	-18.00	-0.18	0.13	0.76	1.47	-0.16	-12.17	0.00	-2.43	2.39	0.00
60	-0.16	-18.40	-0.17	0.13	0.28	0.84	-0.17	-16.15	0.00	-2.51	2.39	0.00

Panel B: using FTSE100 reference benchmark

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.01	-1.48	0.00	0.06	0.22	2.97	-0.01	-1.46	0.16	-2.38	2.50	0.14
24	0.00	-0.39	-0.01	0.08	0.25	3.69	0.00	-0.39	0.69	-2.17	2.35	0.68
36	0.00	-0.62	-0.02	0.11	0.93	4.16	0.00	-0.60	0.53	-2.46	2.19	0.52
48	0.00	-0.42	-0.02	0.14	0.84	2.41	0.00	-0.41	0.68	-2.29	2.15	0.68
60	0.03	3.41	0.02	0.14	0.48	1.23	0.03	3.54	0.00	-2.18	2.26	0.00

The columns show the CARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

Table 6.8: Big–Value Cumulative Abnormal Returns (big-value reference benchmark).

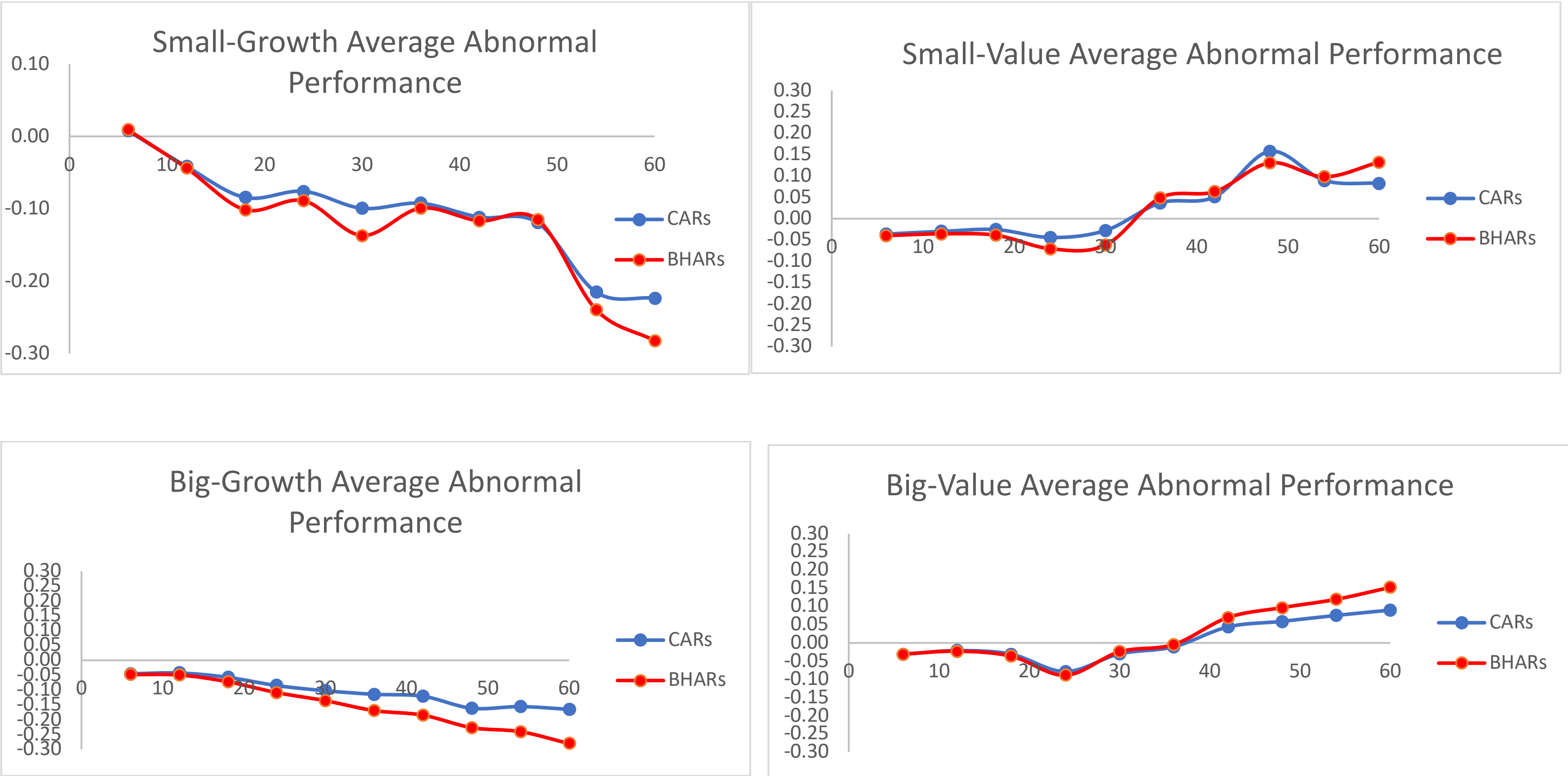
Panel A: using big-value characteristics-based reference portfolio.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.02	-2.25	-0.03	0.08	1.64	4.18	-0.02	-1.89	0.06	-2.22	2.09	0.0
24	-0.08	-4.27	-0.13	0.15	0.96	0.32	-0.06	-3.55	0.00	-2.56	2.35	0.0
36	-0.01	-0.57	-0.04	0.16	0.95	0.93	0.01	-0.54	0.60	-2.51	2.39	0.5
48	0.06	2.28	0.02	0.22	0.98	0.96	0.09	2.50	0.01	-2.41	2.32	0.0
60	0.09	3.79	0.07	0.20	0.37	-0.29	0.12	4.02	0.00	-2.19	2.64	0.0

Panel B: using FTSE100 reference benchmark.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*P-value*	*q-1%*	*q-99%*	*pv1*
12	-0.01	-1.65	-0.02	0.07	0.82	1.24	0.00	-1.55	0.13	-2.26	2.48	0.1
24	0.01	0.68	-0.01	0.09	0.87	0.68	0.02	0.71	0.49	-2.40	2.26	0.5
36	0.02	1.35	0.01	0.11	0.37	-0.22	0.03	1.38	0.16	-2.37	2.25	0.1
48	0.05	2.62	0.04	0.16	0.39	0.08	0.08	2.73	0.01	-2.31	2.63	0.0
60	0.07	3.85	0.06	0.15	0.06	-0.54	0.09	3.89	0.00	-2.50	2.43	0.0

The columns show the CARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 is the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

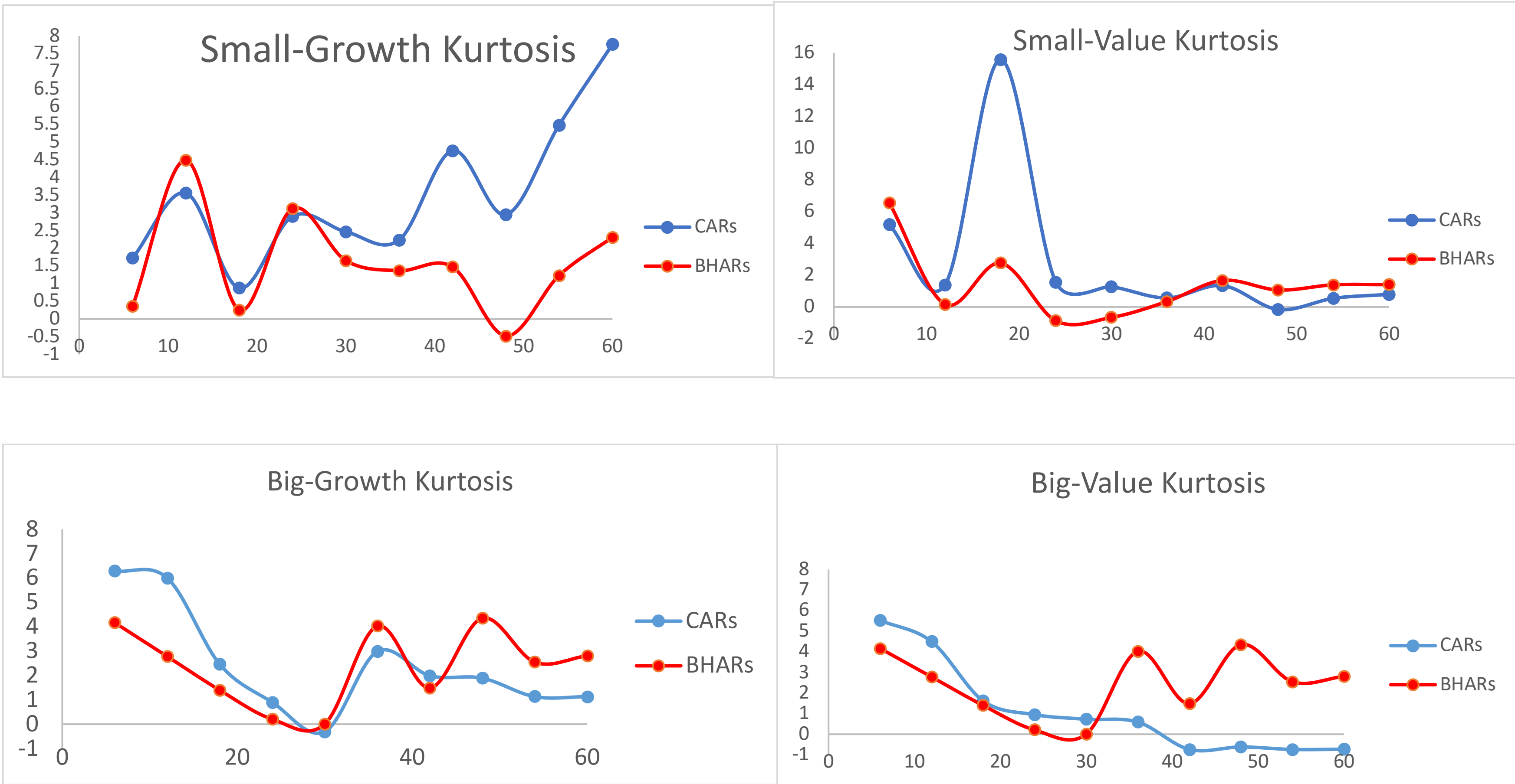


Note: The figure shows scatter diagram of the mean CAR against mean BHAR for four investment styles. Mean excess returns are on the vertical axis and holding period on the horizontal axis.

Figure 6.1: Statistical Properties of the mean CARs and BHARs equally weighted portfolio across four investment styles.

Note: The figure shows scatter diagram of the skewness mean CARs against skewness mean BHARs for four investment styles. Skewness of the mean excess returns are on the vertical axis and holding period on the horizontal axis.

Figure 6.2: Statistical Properties of the Skewness of CARs and BHARs using equally weighted portfolio across four investment styles.



Note: The figure shows scatter diagram of the kurtosis mean CARs against kurtosis mean BHARs for four investment styles. Kurtosis of the mean excess returns are on the vertical axis and holding period on the horizontal axis.

Figure 6.3: Statistical Properties of the Kurtosis of CARs and BHARs using equally weighted portfolio across four investment styles.

6.3. Ethical Funds Results.

Although assets held in ethical funds have more than trebled during the last decade, ethical funds still represent an insignificant proportion of the total mutual funds market. While the increasing popularity of ethical funds investing is welcome news, ethical funds' managers are coming under increasing pressure to improve their financial performance relative to the retail mutual funds' market. In this section, we focus on investigating the long-term performance of UK equity ethical funds, in order to contribute to the debate that exists on ethical fund performance. Like the procedure carried out previously, the BHAR performance is measured according to the hypothesized event of whether ethical funds' style-adjusted performance produces a significant abnormal return over an investment horizon of one- to five years period.

Stylized ethical funds' performance is measured relative to three reference portfolios. The first is a characteristics-based reference portfolio, with factor mimicking portfolios of size and value-growth orientation obtained from Exeter University, are treated as proxies for returns and expected to capture the total risk based on realized return variation. However, ethical funds are subject to ethical constraints and their stock holdings are expected to be different from those of characteristics-based reference portfolios. Second, we considered the stylized ethical funds' performance relative to the UK FTSE4GOOD Index. Thus, we evaluate how stylized ethical funds' returns covary with more general portfolio. Finally, to enhance comparability and capture the effects of ethical criteria on financial performance, we account for the possible return differences between ethical and conventional funds after controlling for fund investment styles. For each stylized ethical portfolio, we selected a characteristic matched portfolio of stylized conventional funds.

This BHAR serves to examine differences in style-adjusted returns between ethical and conventional funds' investment. Thus, abnormal returns are measured based on an investment strategy that could be achieved by systematically buying ethical funds with a specific investment style objective against an equally weighted reference portfolio of its conventional funds' counterparts. Two problems were encountered during this performance comparability. First, ethical funds are less diversified than their conventional counterparts, as they are constructed from a subset of the market portfolio. Second, the number of ethical funds in each stylized portfolio is significantly lower than those of equivalent style controlled conventional portfolios. However, we believe that this is the price of pursuing social objectives, and we implicitly attribute differences in performance to ethical fund screening.

In this section, we attempt to provide evidence on ethical funds' performance using an alternative performance measurement procedure. The BHAR methodology not only captures investors' end wealth, but it also allows for a more detailed ethical fund style analysis. By considering three reference portfolios we aim to provide a robust result of whether ethical funds' investors pay a price for their ethical consideration and whether this penalty differs with fund's different investment styles and horizons.

6.3.1. Ethical Funds BHARs Results

In Table 6.9, Panel A, we report the BHAR of equally weighted small-growth ethical funds across different holding periods. The result reveals a disappointing level of performance of small-growth ethical funds relative to a small-growth characteristics-based reference portfolio. The mean BHAR fluctuate between -6% and -12% over the first four years. Then, the mean returns slump by more than three-fold after five years to reach -37%. All BHARs are highly significant. The median and truncated mean performance closely mimic the performance of the overall mean; the underperformance cannot therefore be attributed to an outlier effect. The mean BHAR distribution is slightly skewed for the first two periods and highly leptokurtic at all investment horizons. When inspecting the skewness-adjusted wild bootstrap test statistics, we found that underperformance is highly significant except for the 24-month holding period. Whilst the kurtosis preserved wild bootstrap test reveals that all conventional test statistics are well-specified, except for the 48-month period, the underperformance is statistically significant at the 5% level. In sum, small-growth ethical funds' investors pay a heavy price for their ethical considerations. The poor performance worsened with time and the exclusion of small-growth stocks based on socially responsible investing criteria has influenced funds' returns negatively.

In Panel B of Table 6.9, we present the results from matching against a market index (FTSE4GOOD). The mean BHAR shows that the performance difference ranges from 4% and -10% for holding periods of 24 and 48 months, respectively. However, this difference is statistically insignificant at any conventional level of significance. The mean BHAR is symmetrical for holding periods over 24 months with high kurtosis at all investment horizons. Furthermore, both the skewness-adjusted and kurtosis preserved wild bootstrap yield similar significance levels to those observed in the conventional t-test at all investment horizons. One might conclude that small-growth ethical funds mimic the performance of FTSE4GOOD more closely than their style-adjusted benchmark (characteristics-based reference portfolio). However, when we compare the dispersion of BHAR using the two reference portfolios, the characteristics-based reference portfolio proved far mor reliable than the FTSE4GOOD index. Particularly, for holding periods longer than 36 months, variance of BHAR using FTSE4GOOD was more than 300 basis points larger than the variance of BHAR using the characteristics-based reference portfolio.

Panel C of Table 6.9 provides a summary of the results for the BHAR using the small-growth conventional funds reference portfolio. The mean BHAR tends to be relatively small in magnitude relative to those obtained with characteristic-based reference portfolio and the FTSE4GOOD. The BHARs fluctuated between -2% and 1% during the first 48-month holding periods, then the rate of decline accelerates with abnormal return reaching -11% after 60-month holding period. The conventional t-test reveal that the difference is only statistically significant at the 5% level for the 1 and 5-year holding periods. Thus, small-growth ethical funds underperformed their conventional counterpart by 2% at the short-term horizon and 11% at the long-term investment horizon. The skewness and the kurtosis of the BHAR is less pronounced than those observed with characteristic-based reference portfolio and the FTSE4GOOD, specifically for holding periods beyond 36 months. Although, the skewness-adjusted wild bootstrap display similar statistical significances to those detected by the standard t-test, the kurtosis preserved wild bootstrap indicates that the performance difference is statistically not different than zero at all investment horizons. Clearly the high kurtosis values affect inferences in BHAR's performance measurement. Thus, failing to account for the high kurtosis values causes rejection of the null hypothesis in short-and long-term investment horizon when it is true. Finally, when it comes to the variance of BHAR, the conventional funds reference portfolio is as good an estimator of expected return as the characteristic-based reference portfolio. The standard deviation ranges from 3% for the 1-year holding period to 19% for a holding period of 5 years. In conclusion, ethical screening neither helps nor hinders small-growth funds' performance over holding periods of 1 to 5 years. Furthermore, by comparing the BHARs result across the three reference portfolios, we can see that ethical funds performed as badly as their conventional fund counterparts. The small-growth characteristic-based reference portfolio beats small growth ethical funds at all investment horizons. Hence, on average ethical funds' managers neither have the skills nor the experience to produce abnormal return that can justify their fees.

Panel A in Table 6.10 shows the mean BHAR results from small-value ethical funds' raw returns against a small-value characteristics-based reference portfolio. There is a steady decline in the equally weighted BHAR, where the BHAR falls from -8% after 12 months to -11% by the end of month 24, and both are statistically significance at the 1% level. Although the rate of decline accelerates to hit -12% after 36 months, -18% after 48 months, and -31% by the end of month 60, none of these abnormal returns is statistically different from zero.

Table 6.9: Small-Growth Ethical Funds Buy and Hold Abnormal Return (BHAR) event time returns.

Holding Period	***Mean***	***t-statistic***	***Median***	***SD***	***Skewness***	***Excess Kurtosis***	***Truncated Mean***	***SK-Adj t-statistic***	***q-1%***	***q-99%***	***pv1***
Panel A: using small-growth characteristics-based reference benchmark.											
12	-0.06	-5.47	-0.06	0.03	-1.00	2.59	-0.06	-8.87	-2.35	2.31	0.00
24	-0.12	-4.74	-0.14	0.08	1.39	2.44	-0.12	-1.20	-2.73	2.09	0.00
36	-0.15	-3.04	-0.14	0.14	-0.83	2.61	-0.15	-3.95	-1.95	2.43	0.00
48	-0.12	-2.46	-0.08	0.14	-0.47	1.12	-0.12	-2.80	-2.21	3.11	0.03
60	-0.37	-5.97	-0.35	0.18	-0.54	0.88	-0.37	-8.16	-2.63	2.21	0.00
Panel B: using FTSE4GOOD reference benchmark											
12	0.00	0.02	0.00	0.03	-1.12	2.92	0.00	-0.04	-2.27	2.33	0.99
24	0.04	1.30	0.01	0.08	1.15	0.82	0.04	1.58	-3.55	2.00	0.21
36	-0.04	-0.74	-0.06	0.17	-0.58	1.09	-0.04	-0.80	-2.84	2.44	0.54
48	-0.10	-1.39	-0.12	0.22	0.78	1.98	-0.10	-1.18	-2.56	2.40	0.24
60	-0.08	-1.04	-0.08	0.22	0.27	1.08	-0.08	-0.99	-2.77	2.63	0.33
Panel C: using small-growth conventional funds reference portfolio.											
12	-0.02	-1.88	-0.02	0.03	-1.20	3.20	-0.02	-2.42	-1.99	2.33	0.11
24	0.01	0.20	-0.02	0.08	1.17	0.99	0.01	0.27	-3.51	2.00	0.91
36	-0.02	-0.39	-0.03	0.15	-0.88	1.68	-0.02	-0.46	-2.36	2.64	0.72
48	-0.01	-0.23	0.01	0.16	-0.77	0.99	-0.01	-0.27	-2.48	2.43	0.87
60	-0.11	-1.73	-0.09	0.19	-0.53	0.87	-0.11	-1.94	-2.64	2.20	0.15

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

As is evident from BHAR variance, the insignificant conventional t-test is attributed to the large level of variation of abnormal returns for holding periods beyond 24 months. Similarly, the median of investment horizons beyond 24 months is significantly worse than the mean. This would suggest that there is a wide variation of average abnormal returns amongst small-value funds given the significant differences between the means and the medians. Thus, there are a few funds managers who produce abnormal returns; however, their performance is hidden by the majority of managers whose performance is poor relative to a small-value characteristics-based reference portfolio. The truncated means display similar results to the overall mean; this can be explained by the low number of small-value ethical funds. For example, only 5 out of 32 ethical funds are tilted towards small-value stocks. The BHAR distribution is fairly symmetrical and highly leptokurtic for investment horizons beyond 12 months. The skewness adjusted wild bootstrap shows a similar statistical power specification to that presented by the conventional t-test. However, the kurtosis preserved wild bootstrap shows that the BHAR performance is highly significant except for the 48 months holding period. Thus, one would conclude that small-value ethical funds exhibit significant underperformance on a style-adjusted returns basis. In effect, small-value ethical fund managers either possess inferior skills which result in shortfalls in performance, or alternatively, strict ethical screening may have limited their performance.

Panel B in Table 6.10 presents the results from a comparison against the FTSE4GOOD index. It appears that the BHAR starts off as being positive for the first and second holding period before tailing off to -1%, -8%, and -4% for the 36, 48, and 60 months, respectively. However, the conventional t-test shows that the difference is statistically insignificant at all investment horizons. The statistical properties of the BHARs' distribution reveal that, the variances are smaller than those obtained with characteristic-based reference portfolio and range from 13% to 23%. Positive skewness and high kurtosis values were observed on the short-and long-term holding periods, but the distribution is close to normal distribution at the medium-term investment horizon. Both the skewness-adjusted bootstrap and the kurtosis preserved bootstrap indicate that the conventional t-test is well-specified. This result would suggest that small value ethical funds' performance closely mimics the performance of the FTSE4GOOD index. However, given the fact that index investing is cost effective in comparison with active investment, small-value ethical funds' investors are expected to generate lower returns relative to the FTSE4GOOD index.

In Table 6.10 panel C, we report the results of the mean BHAR generated by taking a long position in small-value ethical funds and a short position in the small-value conventional funds. This position would have resulted in a loss for investors of 4% to 5% by the end of the second year, 10% by the end of 3 years, 19% by the end 4 years, and 26% by the end of the 5 years in the post style-event period.

The conventional t-tests show that both the 1- and 3-year holding periods are statistically significant at the 10% level, while investment horizons beyond 3 years are highly significant. On the other hand, the skewness-adjusted and kurtosis preserved wild bootstrap shows that the performance difference is only statistically different from zero for holding periods beyond 24 months.

Furthermore, the small-value conventional funds reference portfolio produces the lowest variance of the BHARs, when comparing to the other two reference portfolios. Meanwhile, the BHAR distribution is fairly symmetrical but highly leptokurtic. Thus, on average, ethical funds' managers who effectively follow small-value investment strategies underperform their conventional fund counterparts. One possible explanation for the long-run underperformance is that ethical screening criteria has a negative influence on small-value funds' performance.

Panel A of Table 6.11 presents the results from matching against a big-growth characteristic-based reference portfolio. The mean BHARs is negative and declines exponentially, doubling every year for the first three years. After the first three years, the rate of decline slows, reaching 25% after 4 years and 29% after a 5 year investment horizon. At all investment horizons, the average abnormal return is statistically significant at the 1% level. The conventional t-test is well-specified as is evident from the skewness-adjusted and kurtosis preserved wild bootstrap. The median performance is better than the mean throughout, suggesting that there are a few observations with extremely abnormal returns on the left side of the distribution. Furthermore, beyond the 12-month holding period, the BHARs' distribution is close to normal with skewness excess kurtosis being between 0 and 1. It is worth noting that the BHAR of big-growth ethical funds itself is highly skewed and leptokurtic but match with similar skewness and kurtosis values of the big-growth characteristics-based reference portfolio.

.

Table 6.10: Small-Value Ethical Funds Buy and Hold Abnormal Return (BHAR) event time returns.

Holding Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
Panel A: using small-value characteristics-based reference benchmark.											
12	-0.08	-5.26	-0.08	0.04	-0.76	0.17	-0.08	-8.48	-1.07	1.80	0.00
24	-0.11	-3.71	-0.09	0.07	-1.65	3.10	-0.11	-7.22	-0.88	2.31	0.00
36	-0.12	-1.51	-0.14	0.18	0.31	-1.50	-0.12	-1.38	-1.48	1.17	0.00
48	-0.18	-0.94	-0.32	0.44	0.36	-2.85	-0.18	-0.86	-1.23	1.03	0.40
60	-0.31	-1.62	-0.50	0.43	0.49	-2.79	-0.31	-1.40	-1.31	0.98	0.00
Panel B: using FTSE4GOOD reference benchmark.											
12	0.03	0.53	-0.01	0.13	1.73	3.10	0.03	0.73	-2.34	0.76	0.62
24	0.07	0.78	0.04	0.19	0.84	-0.11	0.07	0.93	-1.80	0.96	0.38
36	-0.01	-0.14	-0.06	0.20	0.91	0.23	-0.01	-0.07	-1.84	1.02	1.00
48	-0.08	-1.03	-0.15	0.18	1.54	2.73	-0.08	-0.67	-2.25	0.92	0.21
60	-0.04	-0.35	-0.10	0.23	1.74	3.26	-0.04	-0.19	-2.35	0.79	0.62
Panel C: using small-value conventional funds reference portfolio.											
12	-0.05	-1.72	-0.07	0.06	1.11	0.58	-0.05	-1.15	-1.93	0.96	0.19
24	-0.04	-1.08	-0.05	0.07	0.32	-1.98	-0.04	-1.01	-1.38	1.18	0.41
36	-0.10	-1.87	-0.09	0.12	-0.23	-2.05	-0.10	-2.01	-1.13	1.38	0.00
48	-0.19	-2.55	-0.11	0.17	-0.60	-1.70	-0.19	-3.17	-1.07	1.53	0.00
60	-0.26	-5.97	-0.26	0.10	-0.11	1.75	-0.26	-6.54	-1.58	1.66	0.00

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

Therefore, based on style-adjusted performance, on average, big-growth ethical funds produce significant underperformance compared to a big-growth characteristics-based reference portfolio, and the poor performance is expected to increase with time.

Panel B of Table 6.11 presents the results from a comparison against the FTSE4GOOD index. The BHAR are significantly smaller in absolute value than those obtained with a big-growth characteristic-based reference portfolio. The difference in performance is statistically insignificant, except for the 24-month holding period, with an associated t-statistic of 2.3. Thus, for a holding period of 24 months, big-growth ethical funds outperformed the FTSE4GOOD index by 3%. Although, the skewness adjusted bootstrap confirms this finding, the kurtosis adjusted wild bootstrap rejected the null hypothesis at lower level of significance (i.e. at the 5% level). The BHAR distribution is negatively skewed and highly leptokurtic for a short- and medium-term investment horizon, but close to normal distribution for an investment horizon beyond 36 months. However, the median is significantly higher than the mean, specifically at the long-term investment horizon. This would suggest that there are very few funds managers who produced a significantly poorer performance relative to the FTSE4GOOD index. Their performance seems to have pulled down the mean more than the median. In conclusion, big-growth ethical funds' investors would enjoy abnormal returns of 3% for a holding period of 2 years. However, this abnormal return is expected to fade away in the long run. Furthermore, investors' relative wealth is highly affected by the choice of which ethical funds to invest in.

Panel C in Table 6.11 presents the results from a comparison against the big-growth conventional funds' reference portfolio. The mean BHAR is marginally small in magnitude and ranges from 0% to -4% for investment horizon of a 1- to 5-year period. In none of the five holding periods is the difference statistically significant at any conventional level of significance. Although, the BHARs' distribution is negatively skewed and highly leptokurtic, the skewness-adjusted and kurtosis preserved wild bootstrap produce a similar statistical power to that observed in the conventional t-test. Furthermore, the variance of the BHARs using the conventional funds' reference portfolio seems to closely match the pattern observed with the characteristic-based reference portfolio, except for over the 60-month holding period. In effect, the set of investment screens that restrict the investment opportunities of big-growth ethical funds has no effect on their financial performance post style event. Stock vetting and screening as applied by big-growth ethical fund managers neither generates valuable information nor yields abnormal performance. Thus, any ethical criteria applied to big-growth funds neither adds nor destroys value in terms of style-adjusted performance.

Our final results, presented in Table 6.12, are derived from equally weighted big-value ethical funds. Panel A reports the results from the use of a characteristics-based reference portfolio. The BHAR falls from an insignificant 2% after 1 year to -13% (highly significant) after 2 years. Then the mean BHAR stabilizes between 2 and 4 years, recording a performance of around -13% (significant at the 1% level). After that, there seems to be some recovery, as the mean BHAR is insignificant at the 5-year investment horizon. The median performance is slightly worse than the overall mean for an investment horizon beyond a 3-year holding period.

Table 6.11. Big-Growth Ethical Funds Buy and Hold Abnormal Return (BHAR) event time returns.

Holding -Period	*Mean*	*t-statistic*	*Median*	*SD*	*Skewness*	*Excess Kurtosis*	*Truncated Mean*	*SK-Adj t-statistic*	*q-1%*	*q-99%*	*pv1*
Panel A: using big-growth characteristics-based reference benchmark.											
12	-0.05	-3.90	-0.04	0.06	1.01	4.15	-0.05	-2.80	-2.22	2.32	0.00
24	-0.12	-4.85	-0.16	0.12	0.82	0.22	-0.13	-3.48	-2.32	2.49	0.00
36	-0.21	-10.01	-0.21	0.10	0.01	0.85	-0.21	-9.92	-2.41	2.42	0.00
48	-0.25	-7.94	-0.21	0.15	-0.68	-0.29	-0.24	-10.91	-2.53	2.65	0.00
60	-0.29	-9.02	-0.25	0.15	-0.57	-0.61	-0.28	-12.25	-2.80	2.60	0.00
Panel B: using FTSE4GOOD reference benchmark.											
12	0.00	-0.39	0.01	0.05	-1.06	1.25	0.00	-0.44	-2.42	2.69	0.71
24	0.03	2.34	0.04	0.06	0.36	1.48	0.03	2.49	-2.69	2.33	0.03
36	-0.01	-0.28	0.01	0.09	-1.30	2.93	0.00	-0.33	-2.52	2.54	0.80
48	-0.01	-0.38	0.04	0.16	-0.86	0.49	0.00	-0.42	-2.67	2.20	0.74
60	0.04	0.94	0.11	0.20	-0.79	0.41	0.05	0.86	-2.35	2.51	0.38
Panel C: using big-growth conventional funds reference portfolio.											
12	-0.01	-1.20	0.00	0.06	-1.76	4.99	-0.01	-1.43	-2.06	2.25	0.25
24	0.00	0.15	0.00	0.07	0.34	0.96	0.00	0.16	-2.55	2.26	0.87
36	-0.02	-1.18	-0.01	0.09	-1.68	5.13	-0.01	-1.40	-2.22	2.40	0.28
48	-0.03	-1.09	-0.01	0.15	-1.47	3.49	-0.02	-1.26	-2.23	2.45	0.32
60	-0.04	-1.02	0.01	0.18	-1.03	1.48	-0.02	-1.13	-2.30	2.56	0.36

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

The BHAR distribution is marginally symmetric and highly leptokurtic for holding period of 2- to 4-year period. Both, the skewness-adjusted and kurtosis preserved bootstrap reveal that the performance is significant between 2 and 4 years. However, the kurtosis preserved wild bootstrap shows that the underperformance for the 3-year holding period is only significant at the 10% level. Thus, one would conclude that, on average, big-value ethical funds underperform their characteristics-based reference portfolio at the medium-term investment horizon.

In Panel B of Table 6.12 we present the results based on the market index (FTSE4GOOD). The difference in BHAR performance between big-value ethical funds and FTSE4GOOD is only statistically significant at the 1% level for the 1- and 4-year investment horizons. On average big-value ethical funds recorded an underperformance of -4% and -14% at the short and long-term investment horizon, respectively. The abnormal returns' distribution is symmetrical and leptokurtic for an investment horizon of up to a 3-year period. When inspecting the skewness-adjusted wild bootstrap test statistics, we found that underperformance is highly significant for the 1-year and statistically significance at the 5% level for 3-and 4-year investment horizons. The kurtosis preserved wild bootstrap test shows a lower level of significance to those founded in the conventional t-test (i.e., at the 10% level).

Panel C in Table 6.12 report the results of the BHAR derived from the big-value conventional funds stylized portfolio. The mean BHAR are negative throughout and ranges from -2% for the 12-month holding period to -12% for the 60-month holding period. However, the conventional t-test is statistically significant for a holding period beyond 36 months at various levels of significance. The median is slightly worse than the mean for a holding period beyond 36 months. The skewness of the big-value ethical funds matches the skewness of their conventional funds' counterpart. The distribution of the ethical funds has a fatter tail than those obtained from conventional funds throughout. The skewness-adjusted wild bootstrap shows that the abnormal return is statistically significant at the 5% level for the 36- holding period. On the other hand, the kurtosis preserved wild bootstrap shows that none of the holding periods is statistically significant at any conventional level. When comparing the variance of the BHAR across the three reference portfolios, the standard deviation of the big-value ethical mean relative to the FTSE4GOOD is lower than those observed relative to a characteristic-based reference portfolio and conventional funds stylized portfolio. Overall, our findings suggest that there is no significant difference in performance between big-value ethical funds and big-value conventional funds.

6.3.2. Summary of Ethical BHAR Results

The aim of this section is to measure and explain the long-run style-adjusted performance of UK equity ethical funds. Results on long-term performance are benchmark dependent and also depend on the length of the investment horizon. Based on the characteristics-based reference portfolio, the study shows that investors pay a heavy price for their ethical considerations and that the performance worsens over time.

Table 6.12: Big-Value Ethical Funds Buy and Hold Abnormal Return (BHAR) event time returns.

Holding Period	***Mean***	***t-statistic***	***Median***	***SD***	***Skewness***	***Excess Kurtosis***	***Truncated Mean***	***SK-Adj t-statistic***	***q-1%***	***q-99%***	***pv1***
Panel A: using big-value characteristics-based reference benchmark.											
12	-0.02	-0.61	-0.01	0.06	-0.55	-0.11	-0.02	-0.67	-2.44	2.86	0.59
24	-0.13	-2.59	-0.13	0.12	0.10	-1.67	-0.13	-2.49	-2.34	2.39	0.00
36	-0.13	-3.74	-0.10	0.08	-0.77	-1.37	-0.13	-5.26	-1.77	4.45	0.06
48	-0.14	-2.66	-0.15	0.12	1.42	2.48	-0.14	-1.19	-2.27	1.79	0.00
60	-0.11	-1.44	-0.13	0.19	0.59	0.39	-0.11	-1.23	-2.61	2.07	0.24
Panel B: using FTSE4GOOD reference benchmark.											
12	-0.04	-2.09	-0.03	0.05	-0.75	-1.11	-0.04	-2.59	-2.08	4.24	0.06
24	0.00	-0.23	-0.01	0.05	0.28	1.71	0.00	-0.21	-2.33	1.93	0.73
36	-0.05	-1.59	-0.03	0.08	-0.57	-1.96	-0.05	-1.82	-2.07	4.31	0.12
48	-0.14	-2.78	-0.16	0.12	0.97	0.46	-0.14	-1.69	-3.22	2.07	0.07
60	-0.08	-1.33	-0.08	0.14	0.26	-0.81	-0.08	-1.25	-2.98	2.79	0.28
Panel C: using big-value conventional funds reference portfolio.											
12	-0.02	-1.53	-0.01	0.03	-0.24	-0.13	-0.02	-1.62	-2.04	3.04	0.18
24	-0.01	-0.88	-0.02	0.04	0.18	-0.95	-0.01	-0.85	-2.92	2.51	0.47
36	-0.05	-1.63	-0.03	0.08	-0.67	-1.26	-0.05	-1.93	-2.06	4.23	0.13
48	-0.11	-2.10	-0.14	0.13	0.50	-1.47	-0.11	-1.76	-4.08	2.22	0.14
60	-0.12	-1.84	-0.13	0.15	0.23	-1.63	-0.12	-1.72	-3.52	2.72	0.12

The columns show the BHARs' mean, t-statistic, median, standard deviation, skewness and kurtosis, a truncated mean based upon winsorising at the 1 % and 99% levels. The table also shows the wild bootstrapped skewness adjusted t-statistic, and pv1 the rejection rate of the kurtosis preserving wild bootstrap as suggested by Davidson et al. (2007) together with cut off values of 1% and 99% quantiles obtained from the kurtosis preserving wild bootstrap.

This result is in line with Gregory and Whittaker (2007) who found that ethical and conventional funds both underperform their benchmark.

Certainly, the exclusion of stocks based on an ethical screening has constrained investors' return optimization and negatively influenced ethical fund performance. Furthermore, investors' return is most likely to deteriorate further with the introduction of management fees and expenses. These results lend support to the underperformance hypothesis, given the fact that the ethical funds investment universe is smaller and restricts portfolio diversification. For example, Hong and Kacperczyk (2005) showed that unethical firms such as those involved in energy, chemicals, alcohol, tobacco and gambling have historically outperformed the market. Underperformance is less pronounced for big-value ethical funds, suggesting that an investor seeking to exploit the big-value premium may experience the least underperformance. However, on the basis of variance comparisons, value funds must be judged inferior to growth funds. Thus, value-oriented ethical funds deliver higher variation in relative returns, which increases investors' realized risk. One possible explanation might be that value-oriented ethical funds are less diversified than growth-oriented ethical funds. For example, Guerard (1997) found that ethical screening is more likely to exclude value-stocks than growth stocks, which in turn leads to a low diversification in value-oriented funds.

From using the ethical index FTSE4GOOD, two clear observations emerge. First, our results show that performance difference appears to be statistically insignificant except for big-value ethical funds, where underperformance is recorded at the short-and long-term investment horizon but with weak statistical significance (i.e.,10%level). The result is inconsistent with other research such as that by Mallin et al. (1995) and Gregory et al. (1997), who reported a significant negative relationship between most ethical funds and the Financial Time All Share Price Index (FTASI). However, they explained their results by the tendency for ethical funds to hold small size firms, with big companies performed substantially better than small companies during their studied period (i.e., 1989 to 1993). It is worth noting that the current practice of screening criteria is positive (i.e., companies that set positive examples in social and environmental issues are included) which allows for the inclusion of big companies and increases the diversification of ethical funds.

As is evident from table 5.12 of the previous chapter (5), most ethical funds are tilted towards big-oriented firms. Thus, our results might suggest that on average ethical fund managers' do not possess any superior skills that would allow them to beat their general index and deliver abnormal performance. A second observation is that the ethical index appears to display a lower return variation with growth oriented ethical funds than those observed with characteristics-based reference portfolio. Thus, it appears that growth-oriented ethical funds implement styles that do not deviate from the ethical index, hence they simply track the FTSE4GOOD despite claiming otherwise.

Finally, our results report ethical fund's performance based on the conventional funds reference portfolio. After controlling for fund's investment styles, we find no evidence of a statistically significant difference in return between ethical and conventional fund returns. One exception is found in the small-value funds; the underperformance can be noted at the long-term horizon, where the difference in returns ranges from -10% for the 3-year holding period to -26% for a holding period of 5-years. This result is partially consistent with other reported studies in the UK ethical funds market, for example Gregory and Whittaker (2007) and Bauer et al. (2005) have both found no statistically significant differences in the performance of most ethical funds as compared to style-adjusted conventional counterparts. However, we suggest two possible explanations for our findings with regards to small-value ethical funds. First, ethical screening is priced, and small-value ethical investors lose out compared to conventional counterparts. Second, unlike Gregory and Whittaker (2007) who matched each ethical fund with 3 non-ethical funds, our matching procedure uses an index of style controlled conventional funds. Thus, the performance differences might be attributed to the presence of a large number of funds in a small-value conventional funds stylized portfolio. By comparing the variance of BHARs across the three reference portfolios, we conclude that conventional funds' reference portfolio is better capable of explaining ethical funds' return than the characteristics-based reference portfolio or the ethical index (FTSE4GOOD). This result is consistent with Bauer et al. (2005) who found that ethical funds are more exposed to conventional funds than their ethical index. From our analysis, it becomes clear that the level of performance depends on the benchmark choices made by researchers and the length of the investment horizons.

In sum, our results are partially in keeping with the hypothesis of no significant cost or benefit to investing in ethical funds. However, it should be noted that small-value ethical funds do perform worse than their conventional counterparts at the long-term investment horizon. Furthermore, compared to conventional funds, one would expect ethical funds to have higher management fees as a result of increased monitoring costs. One of the caveats of our analysis however, is that more data periods covering a wide range of ethical funds might enhance our findings. For example, Hutton et al. (1998) reported abnormal returns for balanced ethical funds (a mix of stocks and bonds). Furthermore, our result is dependent on the appropriate style-adjusted benchmark being employed. Thus, we suggest the use of a different paradigm of risk-return for completeness.

6.4. Conclusion

In this chapter, we have measured the impact of size and value-growth style dimensions on UK-equity funds' performance. Using the alternative approach of event studies, we have examined the propensity of these strategies to generate statistically and economically significant abnormal returns for an investment horizon of 1 to 5 years. The null hypothesis of whether the mean BHAR at holding period $\tau = 12, 24, \dots 60 - \text{months}$ is equal to zero was rejected, when a characteristic-based reference portfolio is used. Our empirical findings show significant inferior performance of value-oriented funds up to 36-months post the hypothesized event. Thereafter, there seems to be some recovery as the mean BHAR turn positive, indicating that investors are getting higher compensation for the risk they are taking. In contrast, growth-oriented funds underperform the characteristic-based reference portfolio and disappoint investors at all investment horizons. These findings can be explained by the severity of the down market during the studied period (i.e., 2008 financial crisis and 2011 sovereign debt crisis). Many literatures (i.e., Lakonishok et al., 1994, Cahine, 2008) have pointed out that value-stocks are likely to generate higher returns than growth-stocks in a Bear Market. Hence one would expect value-oriented fund managers to exploit this phenomenon and deliver a positive abnormal performance.

The empirical evidence of the BHAR performance relative to market benchmark (FTSE100) across style dimensions, highlight no significant differences in returns on the short-and medium-term horizon. However, positive abnormal performance is detected for the value-oriented and big growth funds, but only beyond the 48-month investment horizon. We attribute this unusual finding to window dressing and style rotation activities undertaken by managers to improve ex-post performance. In economic downturns in particular, fund managers' seem to track the FTSE100 as it is easier to justify to investors. However, the average fund manager shows little evidence of stock picking skill that would allow them to recover losses and beat their market index in long-run.

In addition, we show that the general pattern for both the CAR and BHAR performance is consistent in direction across all investment styles. However, the BHAR method systematically magnified performance at all investment horizons. Although, this result seems to support previous findings (i.e., Fama, 1998), there is no evidence of long-lasting underperformance as a result of BHAR compounding problem. Specifically, when positive performance has been realized after a period of negative performance, BHAR returns bounced back at a faster and larger magnitude than CAR.

When it comes to the statistical significance test procedures, both the skewness adjusted, and kurtosis preserved wild bootstrap are in large part consistent with the conventional t-test at all investment horizons. Although we believe that our non-parametric tests were successful in addressing the normality assumption, neither the skewness-adjusted nor the kurtosis preserved wild bootstrap were able to take account of cross-correlation of returns within the sample. Another important observation in our results is that there is a wide variation in abnormal returns and the distributions are slightly skewed to the right and highly leptokurtic. In effect there are a few fund managers in possession of the superior skills that allow them to beat their style specific benchmark. However, their performance is hidden by the majority of managers whose performance is below the mean.

In the second part of this chapter, we turn our attention to UK-equity Ethical funds. Unsurprisingly, the results show that ethical funds' long-term performance is benchmark dependent and it also depends on the length of the investment horizon. Based on the characteristics-based reference portfolio, we show that investors pay a heavy price for their ethical consideration and that the performance of these funds worsens over time. Specifically, the exclusion of stocks based on ethical screening has constrained investors' return optimization and negatively influenced ethical fund performance. However, ethical funds' performance based on the FTSE4GOOD and the conventional funds reference portfolio, show that ethical screening neither helps nor hinders ethical funds' performance for holding periods of 1 to 5 years. One exception is found in small-value funds, where ethical funds exhibit significant under performance on a style-adjusted basis at the long-term horizon. A possible explanation is that small-value ethical funds are less diversified than small-value conventional funds. For example, Guerard (1997) suggested that ethical screening is more likely to exclude value-stocks than growth stocks, which in turn leads to a low diversification in value-oriented ethical funds.

For the purpose of completeness, we also note the results from the use of Cumulative abnormal return (CAR), see appendix for details (Table A6.1 to A6.4). The CAR method indicates similar statistical behaviour with regards to ethical funds' returns for every 12 months up to a 60-month investment horizon. However, the mean CAR returns are smaller than the BHAR in terms of absolute magnitude.

This chapter contributes to the existing literature by assessing the impact of investment style on funds' returns within a context that better resembles investors' end wealth. Following Liu and Strong's (2008) principles, we preserve the buy-and-hold property in our portfolios' construction. The abnormal returns are measured based on an investment strategy that could be achieved by systematically buying units in funds with specific investment style objectives. By employing the event study approach, we accurately measure investors' true returns based on the investment strategy under consideration and draw better statistical inferences.

Chapter 7

The Performance of Ethical and Non-Ethical Funds: Calendar Time Results

7.0. Introduction

The BHAR approach is a cross-sectional test of sample means that relies on the assumption of independence of abnormal returns within the sample. When the independence assumption of event funds' abnormal return is violated, the BHAR suffers from cross-sectional correlation of returns and produces biased test statistics. In particular, the test will tend to reject the null hypothesis when it is true, because of a downward bias estimate of the variance of the cross-sectional abnormal returns. Instead, Fama (1998) and Brav et al. (2000) have strongly recommended the use of a calendar-time (CTAR) approach such as the calendar time three-factor regression. The calendar-time approach uses the mean abnormal time-series of event funds' portfolio returns to eliminate the dependence of returns on cross-sectional analysis. Thus, in an efficient market, the portfolio variance accounts for all cross correlations of abnormal returns, and, hence, the statistical significance of Jensen's alpha is well-specified. Furthermore, the bad model problem is less pronounced in calendar time. Fama (1998) showed that most anomalies have disappeared under the CTAR approach.

Although many studies have advocated the use of the CTAR methodology, some authors advise against it owing to several potential pitfalls. Unlike the BHAR approach, CTAR does not reflect investors' experience. The calendar time portfolio requires monthly rebalancing, since the number of event funds is not equally distributed over the sample period. For example, some funds are added or exit each month from the calendar time portfolio. Lyon et al. (1999) also showed that the regression suffers from residual heteroscedasticity which occurs due to the varying number of funds in the event portfolio composition. Finally, Loughran and Ritter (2000) demonstrated that the CTAR approach has a low power to detect abnormal performance, specifically when managers time the events to exploit mispricing. The CTAR weights each period equally, hence it ignores patterns in the timing of corporate event. Although, many studies have highlighted issues in the CTAR methodology, it is still widely accepted as a robustness check for results obtained using the event-time approach (Mitchell and Stafford 2000, and Brav, et al. 2000).

In the remainder of this chapter, we report in more detail on findings of statistical tests of whether stylized event funds exhibit abnormal return over a 60-month holding period. In particular we compare the relative figures between conventional, ethical, and the whole market in the context of UK equity funds. We construct three different measures of expected returns, namely; a characteristics-based reference portfolio, the market model, and a three-factor model. The statistical inferences are robust to alternative variance estimations, specifically the OLS with White robust standard errors and Gregory et al.'s Feasible GLS techniques.

7.1. Non-Ethical (Conventional) Funds Results.

Panel A of Table 7.1 presents the estimates for Equation (4.13), based on an equal weighted scheme and White (1980) robust standard errors. Using the characteristic-based reference portfolio, growth-oriented funds appear to produce negative alpha estimates, whilst value-oriented funds appear to exhibit positive alpha estimates. However, there is no evidence that value-oriented funds outperformed growth-oriented funds at all investment categories, since the alpha estimates are statistically insignificant at any conventional level of significance. The slopes are highly significant and positive, ranging from 0.62 to 0.89. Thus, the average monthly excess return of the four calendar-time portfolios is well explained by the characteristic-based reference portfolios. The adjusted R-squared is a reasonably high 76.9%, 81.3%, 64.9, and 73.1% for small-growth, small-value, big-growth, and big-value funds, respectively. This suggests a strong relation between systematic risk and expected returns.

Using both the market model and the three-factor model, the intercepts are identical. On average, fund managers appear to deliver positive but insignificant alpha estimates at all investment categories. The beta coefficients on the market portfolio (FTSE100) are higher than those obtained from the characteristic-based reference portfolio except for the big-growth portfolio. The beta coefficients of SMB and HML of the three-factor model are statistically insignificant at any conventional level, suggesting that neither the size-orientation proxied by SMB, nor the growth-value orientation proxied by HML are significant factors in explaining the time-series variation in excess returns of the UK equity funds. The adjusted R-squared remains high and consistent for both the market and three-factor models. However, by comparing the adjusted R-squared across the three models, we can see that the characteristic-based reference portfolios performed better in explaining the cross-section of small-oriented funds returns. Furthermore, the market model and the three-factor model are superior to the characteristic-based reference portfolio for big-oriented fund returns. In general, the results lend support to the single factor model, whether they use the characteristic-based portfolio or a standard market index, given that the intercept is not statistically different from zero, and that the single factor model captures most of the variation in average-returns as well as the three-factor models.

In Panel B, we run similar regressions, but with Gregory et al.'s Feasible GLS variance estimators. The results are similar to those observed in the OLS with a small difference in the coefficient estimates. In particular, the alpha estimates are marginally lower for small-oriented funds, using the characteristics-based reference portfolio model. The t-statistic estimates showed similar pattern to those obtained in the OLS with White robust standard errors. This findings are consistent with Gregory et al. (2010) who report similar standard error estimates for both their version of Feasible GLS and the sandwich variance estimators (White's robust standard errors) with OLS. However, the Feasible GLS does not seem to add explanatory power to the models, for example the adjusted R-squared reveal mixed results. The Feasible GLS demonstrates a higher explanatory power for big-oriented funds, while the OLS with White's robust standard errors shows better fit for small-oriented funds. It is worth noting that the difference in adjusted R-squared across the two techniques is less than 1%.

Our general conclusion is that, on average, conventional fund managers neither added nor destroyed value using both risk-adjusted returns and style-adjusted returns. The intercepts term, α, which represents the abnormal returns of the CTAR portfolios over a 5-year investment horizon, is statistically indistinguishable from zero. However, the performance is likely to deteriorate after accounting for transaction costs and management fees. These results are inconsistent with the performance suggested from the BHAR & CAR methodologies (see Chapter 6).

However, the results are consistent with previous studies such as Lyon et al. (1999), Jegadeesh (2000), and Loughran and Ritter (2000), who argue against using the CTAR methodology since it has low power in detecting abnormal returns and biased toward finding results consistent with market efficiency.

7.2. Ethical Funds Results.

Panel A of Table 7.2. presents the results of four stylized calendar-time ethical portfolios for a 5-year holding period. Within each calendar-time portfolio, ethical funds are treated as equally important and the regression of the characteristic-based reference portfolio, the market model, and the three-factor model are estimated by OLS with White's correction for heteroscedasticity. Using the characteristic-based reference portfolio, the t-statistic estimates shows that alphas are only significantly different from zero for big-growth funds on the style spectrum. Thus, on average, big-growth ethical funds underperformed their characteristic-based reference portfolio by 0.5% per month or 30% over the 5-year holding period. Furthermore, it is clear that value-oriented funds outperformed growth-oriented funds, whereby the magnitude of the underperformance is more pronounced in growth-oriented funds. The slope coefficient (β) is highly significant and positive, recording 0.79 for small-growth, 0.59 for small-value, 0.97 for big growth, and 0.86 for big value. Thus, on average, ethical funds are less risky than their style-adjusted benchmark. The adjusted R-squared range between 58.16% for big-growth and 78.96% for small-value.

Table 7.1: OLS and GLS regression results for equally weighted Non-Ethical Calendar Time stylized portfolios.

		Characteristic-based portfolio				**Market Model**				**3-Factor Model**			
		SG	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***
Panel A. Robust OLS	*Intercept*	-0.002	0.002	-0.003	0.000	0.002	0.002	0.002	0.001	0.002	0.002	0.002	0.001
	t-stat	-1.19	1.04	-1.54	-0.05	0.93	1.14	1.31	0.81	0.90	1.09	1.18	0.78
	β_1	0.76	0.62	0.89	0.74	0.82	0.83	0.74	0.81	0.80	0.82	0.74	0.82
	t-stat	17.41*	13.83*	12.85*	13.60*	13.37*	13.39*	14.75*	13.67*	13.75*	13.77*	15.64*	15.45*
	SMB	-	-	-	-	-	-	-	-	0.11	0.07	0.03	-0.01
	t-stat	-	-	-	-	-	-	-	-	1.46	0.89	0.43	-0.06
	HML	-	-	-	-	-	-	-	-	-0.02	-0.02	-0.04	0.00
	t-stat	-	-	-	-	-	-	-	-	-0.21	-0.23	-0.62	0.05
	Adj-R^2	76.91	81.32	64.97	73.10	75.85	75.90	82.47	80.89	76.51	76.17	82.58	80.90
Panel B. Feasible GLS	*Intercept*	-0.001	0.001	-0.003	0.000	0.002	0.002	0.002	0.001	0.002	0.002	0.002	0.001
	t-stat	-0.82	0.89	-1.61	0.15	1.14	1.35	1.71	0.85	0.99	1.27	1.64	0.82
	β_1	0.77	0.66	0.84	0.75	0.81	0.84	0.80	0.82	0.79	0.82	0.78	0.82
	t-stat	19.90*	22.71*	14.04*	18.14*	18.59*	19.02*	24.93*	22.43*	17.48*	17.58*	23.69*	21.05*
	SMB	-	-	-	-	-	-	-	-	0.12	0.09	0.06	-0.01
	t-stat	-	-	-	-	-	-	-	-	1.84	1.39	1.46	-0.10
	HML	-	-	-	-	-	-	-	-	-0.01	-0.02	-0.03	0.01
	t-stat	-	-	-	-	-	-	-	-	-0.18	-0.22	-0.46	0.07
	Adj-R^2	77.15	81.49	62.70	73.64	74.60	75.50	84.34	81.01	76.00	75.87	84.63	80.93

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression : $R^j_{\tau,t} - rf = \alpha + (R_{\tau,t})^E + \varepsilon_t$, Where: $j = SG, SV, BG, and\ BV$, based on the period Jan 2005 to Jul 2017. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. * Indicates significance at the 1 percent level. ** Indicates significance at the 5 percent level.

An important note of concern is that the regression with the lowest slope coefficient (small-value funds) displays the largest explained proportions of return, while the regression with a slope coefficient close to unity (big-growth funds) shows the largest unexplained proportions of return. There are two possible explanations for this phenomenon. One explanation might be the possible model misspecification in which the characteristic-based reference portfolio fails to capture completely the characteristics relevant for returns (Fama, 1998). Another explanation might be due to the time-varying nature of returns. Mitchell and Stafford (1997) showed that changes in the composition of the event portfolio generate substantial variation in the slope.

Under both the market model and the three-factor model, the abnormal returns are statistically insignificant at any conventional level of significance for a holding period of 60-month. Both models display significant loadings on market risk (FTSE4GOOD) and larger than those observed with a characteristic-based reference portfolio except for big-growth funds. The three-factor model shows insignificant loading on SMB and HML, except for a small-growth calendar-time portfolio, where the SMB factor loading is positive and significant at the 10% level. Both the market model and the three-factor model show a marginally larger adjusted R-squared than those obtained with the characteristic-based reference portfolio, except for small-value funds. By comparing the three models, one would argue that the CAPM explains the cross-section of ethical funds returns relatively well when portfolios are sorted on size and value/growth orientation, especially in the long run.

The results of CTAR portfolios in Table 7.2, Panel B, based on Gregory et al.'s Feasible GLS variance estimators show similarities to the OLS with White's robust standard errors technique. Inspection of the abnormal returns show that the underperformance of big growth funds is ruled out since none of the risk-adjusted performance is statistically significant at any conventional level of significance. The beta coefficients are almost identical to those obtained with the OLS technique. Likewise, the adjusted R-squared figures show a similar pattern to those observed in Panel A of Table 7.2.

In conclusion, ethical fund managers do not in general make excess risk-adjusted returns for their investors. We also specifically observe negative excess risk-adjusted returns in big-growth funds diminishing when moving from estimating the characteristic-based reference portfolio model to the single factor and the three-factor models. These results are inconsistent with previous studies. For example, Gregory and Whittaker (2007) showed that, between 1989 to 2002, UK ethical funds have performed worse than the market benchmark. The results also contradict the BHAR, and CAR results reported in the previous chapter (6).

7.2.1. Ethical versus Conventional Fund Performance.

Next, we turn our attention to comparing the difference in returns between ethical funds' calendar-time portfolios and their conventional funds' counterparts. Similar to the procedure applied before, we used three expected return models to explain ethical funds' monthly conventional funds adjusted returns. The dependent variable is simply the difference between the ethical calendar time return and the conventional calendar time return.

Table 7.2. OLS and GLS regression results for equally weighted Ethical Calendar Time stylized portfolios.

		Characteristic-based portfolio				**Market Model**				**3-Factor Model**			
		SG	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***
Panel A. Robust OLS	*Intercept*	-0.003	-0.001	-0.005	-0.002	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000
	t-stat	-1.36	-0.44	-1.77	-1.00	0.54	-0.21	0.31	-0.21	0.43	-0.20	0.38	-0.18
	β_1	0.79	0.59	0.97	0.86	0.90	0.80	0.84	0.90	0.87	0.78	0.83	0.90
	t-stat	13.16*	13.35*	10.63*	12.89*	14.92*	12.84*	14.20*	11.14*	15.99*	13.39*	15.63*	13.24*
	SMB	-	-	-	-	-	-	-	-	0.14	0.09	0.07	-0.01
	t-stat	-	-	-	-	-	-	-	-	1.73	1.13	0.93	-0.09
	HML	-	-	-	-	-	-	-	-	-0.06	0.01	0.04	0.02
	t-stat	-	-	-	-	-	-	-	-	-0.64	0.08	0.64	0.17
	Adj-R^2	71.92	78.27	59.89	76.81	77.62	75.70	82.11	78.94	78.47	76.16	82.50	78.95
Panel B. Feasible GLS	*Intercept*	-0.003	-0.001	-0.004	-0.002	0.001	-0.001	0.001	0.000	0.001	-0.001	0.001	0.000
	t-stat	-1.28	-0.39	-1.53	-1.04	0.30	-0.45	0.58	-0.21	0.30	-0.33	0.41	-0.18
	β_1	0.80	0.60	0.92	0.86	0.90	0.85	0.83	0.91	0.87	0.81	0.83	0.91
	t-stat	17.40*	21.07*	12.75*	20.27*	20.77*	19.66*	23.10*	20.90*	19.06*	18.03*	21.62*	19.71*
	SMB	-	-	-	-	-	-	-	-	0.13	0.11	0.07	-0.01
	t-stat	-	-	-	-	-	-	-	-	2.11**	1.81	1.27	-0.12
	HML	-	-	-	-	-	-	-	-	-0.06	0.00	0.04	0.01
	t-stat	-	-	-	-	-	-	-	-	-0.72	-0.06	0.61	0.17
	Adj-R^2	71.96	78.96	58.16	77.69	78.48	76.54	81.95	78.74	78.99	76.87	82.49	78.76

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression : $R^{j}_{\tau,t} - rf = \alpha + (R_{\tau,t})^{E} + \varepsilon_t$, Where: $j = SG, SV, BG, and\ BV$, based on the period from Jan 2005 to Jul 2017. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. * Indicates significance at the 1 percent level.

Thus, the factor loadings in each of these regressions represents differences between ethical and conventional funds' average exposure to the factors.

Panel A in Table 7.3. presents the results of the regressions' coefficients, based on OLS with White's (1980) robust standard errors. Under the characteristic-based reference portfolio model, the average abnormal monthly return is small; 0.2 per month or 12% per 5-year holding period across the four stylized calendar-time portfolios. However, the t-statistic estimates show that abnormal returns are only significantly different from zero at the 5% level as funds approach the value-orientation of the style spectrum. Thus, value-oriented ethical funds have underperformed their respective conventional funds by 20 basis points per month over a 5-year investment horizon. This result is partially inconsistent with the underperformance reported in the BHAR and CAR from Chapter 6. Indeed, under the event time methodology, the underperformance is larger in magnitude for small-value ethical funds, while there is no significance, economic or statistical, of the underperformance in big-value ethical funds. Despite the fact that the BHAR method tends to inflate performance because of the compounding effect, the CTAR metric exhibits higher power of the test than an equivalent event time portfolio. The difference in coefficients reveals that only big-oriented ethical funds have significantly different characteristic-based reference portfolio exposure relative to their respective conventional funds. Thus, big-oriented ethical funds have more exposure to the characteristic-based reference portfolio by around 10% (significant at the 1% level). This result is at least partially due to the fact that ethical funds are subject to ethical constraints and their stock holdings are expected to be different from those of a characteristics-based reference portfolio.

The results of the market and three-factor model regressions reveal similar results to those obtained from the BHAR and CAR method. It appears that the performance is significantly negative at the 5% level for small-value ethical funds, whilst the big-oriented funds, together with the small-growth funds, show significantly different market exposure relative to their conventional counterparts. These figures range from 6% for small-growth funds to around 8% for big-oriented funds, over five years at various level of significance. The SMB and HML coefficients of the three-factor model do not offer any extra explanatory power over the market model, except for big-growth funds. The big-growth ethical funds recorded a positive HML coefficient of 9% statistically significant at the 5% level. This suggests that value-growth effects played a role in the covariance differences between big-growth ethical funds and their respective conventional funds. The positive loading on HML indicates that big-growth ethical funds are less exposed to growth stocks relative to their conventional funds.

From Table 7.3, Panel B, the results based on Gregory et al.'s Feasible GLS variance estimators, show that there is no significant difference in the mean performances with the OLS technique. Although the Feasible GLS generates slightly higher t-statistics than those obtained under OLS, the significance of the underperformance is only strengthened for a small-value calendar time portfolio using the market model regression. For example, underperformance becomes statistically significant at the 1% level. However, with regards to the beta coefficients, the Feasible GLS technique offers no pattern that would allow it to be favored over the OLS with White's (1980) corrected standard errors.

In short, viewed from the perspectives of both the Feasible GLS and the OLS with White's (1980) techniques, ethical funds underperformed their conventional counterparts significantly as funds approach the value-orientation of the style spectrum. This result is partially in keeping with the Gregory et al. (1997), Bauer et al. (2005), and Gregory et al. (2007) who showed that ethical funds did not perform as well as their conventional funds under time invariant risk factors (static model). Our evidence further suggests that there are significant differences in factor exposure between ethical funds and their conventional counterparts. The differences in beta are more pronounced for big-oriented funds under the three models of expected returns. Thus, big-oriented ethical funds have a higher market exposure than big-oriented conventional funds. This result is inconsistent with those of previous UK based studies (i.e., Gregory et al. (2007), who showed that the direction of the differences tends to be toward higher expected returns, with ethical funds having a higher exposure to small and value stocks than conventional funds, in an attempt to beat the market.

7.2.2. Aggregate Ethical versus Conventional Funds: five-year holding period

Table 7.4. reports the results of an equally weighted calendar time portfolio of the overall or collective ethical funds in the sample and those of equally weighted conventional funds. The returns of the ethical/ conventional calendar-time portfolio can be achieved by systematically buying units in funds with ethical/ conventional investment criteria over an investment horizon of a five-year period. The table presents the performance of the two samples with the expected returns generated by the market model and the three-factor model.

Table 7.3: OLS and GLS regression results for the difference between the Ethical Calendar Time stylized portfolios and the Conventional Calendar Time stylized portfolios.

		Characteristic-based portfolio				Market Model				3-Factor Model			
		SG	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***	***SG***	***SV***	***BG***	***BV***
Panel A. Robust OLS	*Intercept*	-0.001	-0.002	-0.002	-0.002	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002
	t-stat	-0.60	-	-1.41	-	-0.66	-	-1.31	-1.47	-0.73	-	-1.01	-1.38
	β_1	0.03	-0.03	0.09	0.11	0.06	-0.05	0.09	0.08	0.06	-0.05	0.07	0.08
	t-stat	1.11	-1.56	2.43*	4.04*	2.25**	-1.80	2.78*	1.98	1.87	-	2.36*	2.12**
	SMB	-	-	-	-	-	-	-	-	0.03	0.02	0.04	0.00
	t-stat	-	-	-	-	-	-	-	-	0.87	0.55	1.37	-0.03
	HML	-	-	-	-	-	-	-	-	-0.03	0.03	0.09	0.02
	t-stat	-	-	-	-	-	-	-	-	-0.70	0.79	2.54*	0.34
	Adj-R^2	1.77	2.97	6.04	18.12	5.05	3.13	11.61	8.51	5.84	4.09	18.07	8.64
Panel B. Feasible GLS	*Intercept*	-0.001	-0.002	-0.002	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001
	t-stat	-1.01	-	-1.52	-	-0.84	-2.32*	-1.12	-1.52	-0.94	-	-0.72	-1.50
	β_1	0.01	-0.03	0.09	0.11	0.04	-0.05	0.08	0.07	0.03	-0.05	0.06	0.07
	t-stat	0.66	-1.93	2.81*	4.94*	1.87	-1.92	3.70*	3.15*	1.31	-	2.02**	2.95*
	SMB	-	-	-	-	-	-	-	-	0.03	0.02	0.05	0.00
	t-stat	-	-	-	-	-	-	-	-	0.95	0.55	1.47	-0.01
	HML	-	-	-	-	-	-	-	-	-0.01	0.04	0.09	-0.01
	t-stat	-	-	-	-	-	-	-	-	-0.26	0.82	2.23**	-0.14
	Adj-R^2	0.80	3.08	6.04	17.07	3.07	3.13	10.37	7.63	3.50	3.74	14.59	7.55

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression : $ER^j_{\tau,t} - CR^j_{\tau,t} = \alpha + (R_{\tau,t})^E + \varepsilon_t$, Where: $ER^j_{\tau,t}$ is the equally weighted calendar-time ethical funds' portfolio in month t for a holding period $\tau = 60 - month$. $CR^j_{\tau,t}$ is the equally weighted calendar-time conventional funds' portfolio in month t for a holding period $\tau = 60 - month$. $j = SG, SV, BG, and\ BV$, based on the period from Jan 2005 to Jul 2017. The symbols used to denote the investment style imply the following: SG = small-growth, SV=small-value, BG=big-growth, and BV=big-value mimicking portfolios. *Indicates significance at the 1 percent level. ** Indicates significance at the 5 percent level.

Our intention is to explicitly identify and report on the performance of ethical and conventional funds over the whole sample, with no specific style consideration. As a first sign of risk-adjusted performance, in Panel A of Table 7.4. we present the OLS estimates for Equation (4.13) with White's (1980) robust standard errors. Over a five-year holding period, the intercept of both CAPM and the three-factor model show that neither ethical nor conventional funds exhibit a significant raw return performance. The betas on the market portfolio are highly significant across both factor models. Furthermore, the slope coefficients of SMB and HML are both statistically insignificant at any conventional level. The adjusted R-squared is high and consistent at both the market and the three-factor model. Thus, around 80% of the variation in excess returns of ethical and conventional calendar time portfolio can be explained by variation in market returns (FTSE4GOOD/FTSE100). These results are largely consistent with the findings reported in the literature (see for example Gregory et al., 1997; and Bauer et al. 2005.

The results in Panel B of Table 7.4. based on Gregory et al.'s Feasible GLS indicate similar results to those obtained with the OLS and White's robust standard errors. Although the magnitude of the performance is not different from those observed in Panel A, there is an improvement in the confidence level observed under the Feasible GLS.

Panel A Table 7.5. presents the OLS regression results of the performance difference between the returns on ethical funds and the conventional calendar time portfolio. For both the single-and the three-factor model, on average ethical investors lose out to non-ethical investors by around -0.1% per month or 6% over a 5-year period. The t-statistic estimates show that the underperformance is only significantly different from zero at the 10% level.

Table 7.4: Regression coefficients for equally weighted of ethical and conventional calendar time portfolio on two benchmark models.

		Market Model		***3-Factor Model***	
		Ethical	***Conventional***	***Ethical***	***Conventional***
Panel A. Robust OLS	*Intercept*	0.000	0.002	0.000	0.002
	t-stat	0.11	1.08	0.10	1.02
	β_1	0.86	0.80	0.84	0.79
	t-stat	13.72*	14.31*	15.31*	15.21*
	SMB	-	-	0.07	0.05
	t-stat	-	-	0.88	0.69
	HML	-	-	0.00	-0.02
	t-stat	-	-	0.02	-0.24
	Adj-R^2	81.45	80.20	81.72	80.36
Panel B. Feasible GLS	*Intercept*	0.000	0.002	0.000	0.002
	t-stat	-0.05	1.34	-0.04	1.29
	β_1	0.88	0.84	0.86	0.82
	t-stat	22.76*	21.83*	21.15*	20.66*
	SMB	-	-	0.07	0.08
	t-stat	-	-	1.34	1.54
	HML	-	-	0.00	-0.03
	t-stat	-	-	-0.03	-0.39
	Adj-R2	81.41	80.41	81.71	80.82

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression : $R_{\tau,t} = \alpha + (R_{\tau,t})^E + \varepsilon_t$, Where: $R_{\tau,t}$ is the equally weighted calendar-time portfolio in month t for a holding period $\tau = 60$, based on the period from Feb 2005 to Jul 2017.
*Indicates significance at the 1 percent level. ** Indicates significance at the 5 percent level.

This result is in keeping with other studies (i.e., Gregory et al. (1997) and Kreander et al. (2005)) who documented a similar performance pattern. Although they claimed that the performance difference is due to the size effect, our results show that ethical funds have significantly greater exposure to a market portfolio by around 4%, significant at the 1% level using the CAPM, and at the 5% level using the three-factor model. It is worth noting that, during the sample period of Gregory et al. (1997) and Kreander et al. (2005), ethical funds operated on negative screening practices, which resulted in the exclusion of big stocks.

Gregory et al.'s Feasible GLS model in Panel B of Table 7.5. provides better inferences than the OLS with White's robust standard errors. The statistical significance of ethical funds underperformance is even more marked under the GLS model. For example, the CAPM shows that the underperformance is statically significant at the 5% compared to 10% in the OLS estimation, while the three-factor model shows a similar statistically significant level.

7.3. Conclusion

In this Chapter, we carried out the calendar-time approach to detect long-term abnormal performance of the UK equity funds over a 5-year holding period. By distinguishing between conventional and ethical funds, we explored whether funds' performance differed across styles and scrutinised their ability to generate abnormal returns. We also compared the performance of ethical funds relative to their conventional peers with the same investment style and investigated what incentives might exist to explain these results. The calendar-time portfolio is equivalent to an investment strategy that could be achieved by investing a fixed amount of cash in mutual fund portfolios over an investment horizon of a five-year period.

Table 7.5: Regression coefficients for the difference between the Ethical Calendar Time portfolios and the Conventional Calendar Time portfolio.

		Market Model	***3-Factor Model***
		Ethical Vs Conventional	***Ethical Vs Conventional***
Panel A. Robust OLS	*Intercept*	-0.001	-0.001
	t-stat	-1.83	-1.67
	β_1	0.05	0.04
	t-stat	2.01**	1.69
	SMB	-	0.02
	t-stat	-	0.87
	HML	-	0.03
	t-stat	-	0.86
	Adj-R^2	5.59	7.25
Panel B. Feasible GLS	*Intercept*	-0.001	-0.001
	t-stat	-1.97**	-1.88
	β_1	0.04	0.04
	t-stat	2.59*	2.14**
	SMB	-	0.02
	t-stat	-	0.74
	HML	-	0.03
	t-stat	-	0.79
	Adj-R^2	5.37	6.57

The table reports the intercepts, slope coefficients and adjusted R-squared of the regression : $ER_{\tau,t} - CR_{\tau,t} = \alpha + (R_{\tau,t})^E + \varepsilon_t$, where: $ER_{\tau,t}$ is the equally weighted ethical calendar-time portfolio in month t for a holding period $\tau = 60$, $CR_{\tau,t}$ is the equally weighted conventional calendar-time portfolio in month t for a holding period $\tau = 60$, based on the period from Feb 2005 to Jul 2017. *Indicates significance at the 1 percent level. ** Indicates significance at the 5 percent level

Although the weighting schemes may play an important role when detecting long-run abnormal returns, data on funds' size are not easily attainable in the UK mutual fund industry. Another problem we faced is that the long-term abnormal return is sensitive to the choice of expected returns model. Accordingly, we used the characteristic-based reference portfolio along with the market model and the three-factor model to capture long term expected returns. We use White's robust standard errors and Gregory et al.'s Feasible GLS techniques to deal with heteroskedasticity caused by different number of funds in each calendar month. Our empirical results are threefold. First, in relation to both ethical and conventional calendar time portfolios, under the three expected returns models, we find no evidence of abnormal performance when funds are sorted on a style-adjusted basis. Although we documented negative mean intercepts under the characteristic-based reference portfolio model, the performance is statistically insignificant at any conventional level of significance

These results are inconsistent with the performance suggested from the BHAR and CAR methods in Chapter 6. However, the results are consistent with the finding of Cuthbertson et al. (2008) who show that fund managers are not able to beat their benchmarks after controlling for size, and value/growth factors. Our results seem to be supportive of those in Lyon et al. (1999), Jegadeesh (2000), and Loughran and Ritter (2000) who argue against using the CTAR method as it has low power to capture abnormal returns and biased toward finding results consistent with market efficiency. Second, we find no significant difference between the OLS with White's corrected standard errors and Gregory et al. (2010) Feasible GLS variance estimators. However, the GLS model offers slightly better statistical inferences and fit in term of adjusted R-squared. One possible explanation might be that the number of ethical funds in the calendar time portfolio is particularly low, and ranges from 19 to 24 for each calendar month. Although White's corrected standard error is useful in the presence of unknown form of heteroscedasticity, MacKinnon and White (1985) demonstrated the risk of false inferences when the sample size is small. Overall, our finding is consistent with Gregory et al. (2010) who report broadly similar standard error estimates for both techniques.

Finally, by comparing the risk-adjusted performance of ethical funds relative to their conventional peers, the results showed that the intercepts are negative and significantly different from zero as funds approach the value-orientation of the style spectrum. These results are partially corroborated with previous research on ethical fund performance (Gregory et al., 1997; Bauer et al., 2005; and Gregory et al., 2007). Investing in value-oriented ethical funds does lead to returns that are significantly lower than those delivered by conventional funds. Furthermore, our evidence suggests that there are significant differences in market exposure between ethical funds and their conventional peers. The direction of the differences tends to be toward lower expected return. Thus, ethical funds' performance is consistent with the goal of matching the market proxy, rather than beating the market. In addition, the performance difference is likely to widen further after accounting for transaction cost and management fees. Bauer et al. (2005) report that ethical funds are typically smaller in size, and more likely to charge a higher expense ratio. Therefore our results lend support to the claim that imposing ethical constraints leads to a weaker investment performance.

However, the validity of our findings is based on the assumption that the model of expected returns is correct. Overall, our results are in favour of the characteristic-based reference portfolio model with Feasible GLS robust standard errors. The characteristic-based reference portfolio model not only has theoretical rationale, but it also resonates well in practice, as it captures most of the variation in average-returns and has a higher power to detect long-term abnormal performance. Although our calendar-time portfolios were constructed under equally weighted scheme, the model reasonably captures the variation in small-oriented funds. This is in contrast to Fama (1998) who indicated that bad-model problems are more severe for small stocks when inferences are made from equally weighted returns. Nevertheless, the doubts about the contribution of the bad model problems to our results remain an open question.

Chapter 8

Skill versus Luck in Fund Performance

8.0. Introduction

In this chapter, we conduct an examination on the ex-post performance of UK equity funds. We explicitly control for the luck factor, while allowing for the role of the funds' investment style. The question we try to answer is whether some fund managers possess superior/inferior stock-picking skills over their pairs? In other words, are the significant alpha estimates simply the result of the extraordinarily good/bad luck of a few individual fund managers? In contrast to previous studies, which use standard statistical measures, we apply bootstrap procedures across all UK equity funds. We compare the joint distribution of the mutual funds' t-statistics of the alpha obtained from asset pricing models against the simulated luck distribution. A significant difference between these bootstrapped statistics is regarded as evidence of genuine good/bad skills. Two bootstrap procedures are employed, namely, the 'baseline', and the 'skewness-adjusted and kurtosis preserving wild' bootstraps. We examine the empirical results of the two bootstraps based on the gross and net returns of the single and three- factor models, we also investigate whether managers' skill levels differ across different investment styles. These approaches can lead to correct inferences when the distribution of abnormal return is highly non-normal and heteroscedastic. It also allows us to assess whether superior performance is a result of stock picking skills, or simply due to luck.

8.1. Empirical Results: Gross Returns

8.1.1. Baseline bootstrap Results

Table 8.1 displays the baseline bootstrap results for the full sample of UK-equity funds between 2002-2017. In order to assess whether fund performance is superior/inferior to random sampling variation under the null hypothesis of zero alpha, the ranked t-statistic of alpha bootstrap is compared to the actual ranked t-statistics of alpha. Then, the bootstrap p-value is used to draw inferences about managerial stock selection skills at different quantiles of the performance distribution.

Panel A of Table 8.1 reports the ranked performance statistics of the cross-sectional distribution of the actual funds' returns at selected percentiles of the distribution. Row 2 shows that the worst performing fund (Min) based on the three-factor model achieved -2.5% per month, but it is statistically insignificant at the conventional levels (row 3). In row 4, we can see that the lowest ranked t-statistics of actual alpha is -0.937. However, under the imposed null hypothesis of zero alpha, the bootstrap p-value (of the t-statistics of alpha bootstrap $t_{\alpha\ boot}$) in row 5 is 0.817. Thus, there are 81.7% of the lowest 1000 simulations across all funds with a $t_{\alpha\ boot}$ value of less than -0.937. Therefore, we cannot reject the null hypothesis that the performance of the worst fund is due to random sampling variation in the performance estimator around a true value of zero. Hence, the managers of the worst performing fund do not possess poor stock selection skills, and their performance can be explained by bad luck. By following the same assessment across the entire left tail of the distribution (i.e., up to the 50th percentile). We can see that the percentage of the 1000 simulation runs that produced a lower value of $t_{\alpha\ boot}$ than the actual t-statistics of alpha exceeds the significance cut off point of conventional levels of significance. Thus, the performance of the entire left tail is attributed to chance, and there is no indication of poor stock selection skills among funds' managers.

Similarly, looking at the right tail of the distribution, the actual funds' performance can be explained by random sampling variation in the t-statistics around a true value of zero, except for the 95th percentile. For example, the top performing fund achieved abnormal return of 2% per month and it is statistically significant at the 1% level. Although the top ranked t-statistics of actual alpha is 3.78, the bootstrap p-value of the $t_{\alpha\ boot}$ indicates that, 13.8% of the $t_{\alpha\ boot}$ of the top 1000 simulations were higher than 3.78. Hence, we conclude that the top performing fund's managers do not possess superior skills that allowed them to beat the market and achieve abnormal return. The importance of this finding is that inferences based on the parametric test which relies on the normality assumption, would have wrongly concluded that the top ranked t-statistics of the actual fund's alpha is statistically significant, while in fact their performance is simply due to random sampling variation. In contrast, there is strong evidence of genuine stock selection skill within the top 5% ranked funds (95th percentile) significant at the 5% level. The bootstrap p-value indicates that only 2% of the 1000 $t_{\alpha\ boot}$ of the top 5% of ranked funds have a t-statistics higher than 3.005. It is important to note that our baseline bootstrap distribution uses the t-statistic rather than alpha. Hence the highest ranked t-statistic might not correspond to the highest ranked alpha.

In Panel B of Table 8.1, we present the baseline bootstrap analysis under the single factor model. The results show that the actual fund ranked alphas are similar to those observed under the three-factor model. The corresponding p-value is only significant in the top performing fund and the 99th percentile, at significance level below 5%. Inferences from the bootstrap p-value of the $t_{\alpha\ boot}$ show that the performance of the left tail of the distribution can be explained by chance alone. The p-value of the $t_{\alpha\ boot}$ indicates that there are 77.3% of the lowest 1000 simulations across all funds with a $t_{\alpha\ boot}$ value of less than -1.050. However, performance at the 60th, 70th, and 80th percentiles becomes very unlikely to be explained by random sampling variation. It can be interpreted as evidence of superior skills at the 1% level of significance. Thus, fund managers certainly add value to their funds, whether by using private information or genuine stock selection skills. Now, the further we move to the right, for example from the 90th percentile to the extreme tail of the distribution, fund managers become unable to beat their luck distribution at a conventional level of significance.

The important implication of these findings is that, ranking funds according to their performance give us little information about funds managers' skills and the fund's future performance. Retail investors should be aware of such details prior to entering an investment in an actively managed fund, especially given that past performance is commonly used in the marketing of mutual funds.

8.1.2. Wild adjusted bootstrap results

Panel A of Table 8.2 presents the wild-adjusted bootstrap results based on the t-statistics of the three-factor model at selected quantiles of the performance distribution. By comparing the p-value of the actual alpha found in Table 8.1 (row 3) and its counterpart in Table 8.2 (row 3), we find that Johnson's (1978) skewness-adjusted t-statistics has no impact on the significance level of the performance.

Table 8.1: Statistical Significance of UK Equity Fund Performance

Quantile	Min	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	Max
Panel A: Baseline bootstrap statistics under three-factor model															
Alpha	-0.025	-0.008	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.012	0.020
p-value-alpha	0.355	0.188	0.338	0.215	0.735	0.892	0.945	0.242	0.765	0.382	0.281	0.305	0.003	0.044	0.000
t-stat	-0.937	-1.327	-0.965	-1.243	-0.340	-0.136	0.069	1.175	0.300	0.877	1.081	1.036	3.005	2.025	3.780
p-value-t-boot	0.817	0.834	0.860	0.477	0.970	0.995	1.000	1.000	0.428	0.147	0.332	0.685	0.020	0.389	0.138

Panel B: Baseline bootstrap statistics under single factor model															
Alpha	-0.028	-0.008	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.002	0.002	0.003	0.005	0.006	0.013	0.019
p-value-alpha	0.300	0.198	0.193	0.284	0.732	0.960	0.911	0.740	0.401	0.134	0.090	0.058	0.133	0.039	0.000
t-stat	-1.050	-1.297	-1.317	-1.074	-0.344	-0.051	0.112	0.334	0.842	1.506	1.715	1.907	1.512	2.077	3.692
p-value-t-boot	0.773	0.871	0.616	0.657	0.974	1.000	1.000	0.998	0.000	0.002	0.026	0.125	0.491	0.352	0.229

The table shows baseline bootstrap statistics at selected percentiles of the distribution of the three-factor and single-factor model for all UK-equity gross fund returns over the sample period of 2002 to 2017. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) bootstrap's t-statistic estimate than the actual fund's t-statistic estimate.

Although higher t-statistics were observed under Johnson's approach at the extreme tails of the distribution, the significance level (p-value) is very similar across the two performance distributions. This might suggest that the skewness of the actual funds' distribution has little impact on our inference. Turning our attention to the results of the wild-adjusted bootstrap p-value, we find that up to the 90th percentile, the performance can be explained by random variation around a true value of zero. This pattern is similar to those observed in Table 8.1. However, beyond the 90th percentile, the p-values present evidence of significant superior skills at a conventional level of significance. Hence, fund managers at the extreme right tail of the performance distribution were able to beat their luck distribution at a conventional level of significance. For example, the top 5% ranked funds show an actual ex-post t-statistic of 2.98 with a bootstrap p-value of 0.3%. This indicates that only 0.3% of the bootstrapped t-statistics from 1000 simulations across all funds were higher than 2.98 under the null hypothesis of zero alpha. When we operate at a 5% upper tail cut off point, we can reject the hypothesis that the performance of the top 5% funds is purely because of luck at the 5% level. By comparing these results with those obtained from Table 8.1, we can see that the wild adjusted bootstrap provides improved inferences in distinguishing skill from luck in fund performance. Thus, one would conclude that, beyond the 95th percentile, the baseline bootstrap t-statistics exhibit higher variance and greater non-normality than the wild adjusted bootstrap.

Panel B of Table 8.2 presents the wild-adjusted bootstrap result under the single factor model. In the left tail of the distribution, the bootstrap p-values suggest that the inferior performance can be explain by bad luck rather than poor stock selection skills. The bootstrap p-value of the bottom (Min) performing fund indicates that 75.2% of the

$t_{\alpha\ boot}$ have a t-statistic lower than -1.543. Skilful performance is documented at the 60^{th}, 70^{th}, and 80^{th} percentile, statistically significant at the 1% level. Furthermore, funds ranked at the 99^{th} percentile and the top ranked fund (Max) are also skilful at conventional levels of significance. When comparing the result with its baseline bootstrap counterpart in panel B of table 8.1, we find that the results are almost identical except for the extreme right tail of the distribution. The baseline bootstrap seems to attribute superior performance to chance more often than the wild bootstrap, particularly at and beyond the 99^{th} percentile of the performance distribution.

Overall, the patterns that can be identified from Table 8.1 and Table 8.2 are that, regardless of the model used, there is no evidence of the presence of bad stock selection skills among UK equity fund managers. Our results suggest that the performance of the worst managers is solely due to bad luck. There are very few funds that display stock picking ability in the right tail of the distribution, though the skilful performance is more pronounced under the single factor model. With respect to the bootstrap scheme, both bootstraps show broadly similar results. However, the wild bootstrap is found to be more powerful at the extreme right tail of the performance distribution.

8.2. Empirical Results: Net Returns

8.2.1. Baseline bootstrap Results

Up until now, the Funds' performance is evaluated on a gross returns basis rather than on net returns to investors. Since investors are interested in the end value of their investment after fees and costs have been taken into account, we conduct the analysis for both bootstraps based on the assumption of 1.95% annual fees across all funds.

Table 8.2: Statistical Significance of UK Equity Fund Performance

Quantile	Min	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	Max
Panel A: Wild adjusted bootstrap statistics under three-factor model															
Alpha	-0.025	-0.008	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.012	0.020
p-value-alpha	0.188	0.193	0.330	0.204	0.720	0.889	0.930	0.243	0.778	0.408	0.278	0.318	0.003	0.005	0.000
t-stat	-1.343	-1.312	-0.982	-1.274	-0.360	-0.139	0.088	1.171	0.283	0.829	1.088	1.010	2.978	2.829	5.172
p-value-t-boot	0.789	0.661	0.768	0.411	0.952	0.997	0.999	1.000	0.346	0.070	0.164	0.632	0.003	0.056	0.000

Panel B: Wild adjusted bootstrap statistics under single factor model															
Alpha	-0.028	-0.008	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.002	0.002	0.003	0.005	0.006	0.013	0.019
p-value-alpha	0.131	0.205	0.199	0.271	0.712	0.954	0.924	0.758	0.415	0.134	0.106	0.072	0.127	0.004	0.000
t-stat	-1.543	-1.277	-1.299	-1.104	-0.371	-0.058	0.096	0.309	0.817	1.509	1.633	1.810	1.535	2.934	5.039
p-value-t-boot	0.752	0.705	0.552	0.546	0.945	1.000	0.999	1.000	0.000	0.000	0.012	0.091	0.421	0.057	0.000

The table shows wild adjusted bootstrap statistics at selected percentiles of the distribution of the three-factor and single-factor model for all UK-equity gross fund returns over the sample period of 2002 to 2017. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) bootstrap's t-statistic estimate than the actual fund's t-statistic estimate.

Thus, we computed the net returns for each fund by deducting the monthly equivalent of the assumed annual fund management fee (around 0.163% per month). The bootstrap results will therefore indicate whether fund managers have sufficient skills that allow them to earn abnormal returns for their investors after accounting for operating and management fees. Panel A of Table 8.3 presents the ex-post performance based on net returns under the three-factor model with a baseline bootstrapped p-value at selected percentiles. The results show that underperformance stretched to lengths approaching the 50^{th} percentile, hence under the three-factor model, 50% of our sample funds generated negative net realized returns to investors. Funds' performance ranged from -2.7% per month for the worst performing to 1.8% per month for the top performing fund. At all percentile points except the top performing fund, the p-value of the parametric t-test indicates that the performance is statistically insignificant at the conventional levels. The p-values of the baseline bootstrap show that the performance of funds ranked at the extreme left tail of the performance distribution can be attributed to chance alone. The further we move towards the centre, performance becomes increasingly unlikely to be explained by random sampling variation. The bootstrap p-value of the 5^{th}, 10^{th}, 40^{th}, and 50^{th} percentile is below the 5% cut-off point and can be interpreted as poor managerial skills, where the t-statistics of poor performance exceed the performance which could be explained by bad luck at the 5 % level. The bootstrap p-value for funds ranked above the 50^{th} percentile have largely exceeded the significance cut off point of conventional levels of significance. Thus, funds' performance is attributed to random variation around a true value of zero, and the apparent winners are simply lucky.

Panel B of Table 8.3 presents the abnormal performance based on net returns using the single-factor model along with baseline bootstrap results. The p-value of the conventional t-test shows that the performance is statistically insignificant except for the top performing fund. However, the top ranked t-statistics of actual alpha is 3.38. The bootstrap p-value of the t-statistics of alpha is 0.251 and indicates that 25.1% of the top 1000 simulations have a t-statistic higher than 3.38. Thus, the positive abnormal returns that have been documented in the top performing fund are simply due to chance alone. Looking at both extreme ends of the performance distribution, the bootstrap p-value of the t-statistics shows that the under/overperformance can be simply explained by random variation around the true value of zero alpha. Therefore, fund's managers are found to exhibit bad/good luck rather than bad/good skills. However, there are a small group of unskilful fund managers around the centre of the performance distribution. In particular, the bootstrap p-value of the 20^{th}, 40^{th} and 50^{th} percentile is very close to zero suggesting poor stock selection ability at a 1% level of significance. However, the residual variance (unexplained error) of fund regression tends to be close to zero around the centre of the performance distribution. Therefore, the 1000 simulations of the bootstrap coefficient estimates are more likely to have low sampling variation and to produce a low t-statistic. Hence, the bootstrap methodology is most reliable at the extreme tails of the performance distribution when the variance of fund regression residuals is larger.

Table 8.3: Statistical Significance of UK Equity Fund Performance

Quantile	Min	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	Max
Panel A: Baseline bootstrap statistics under three-factor model (NET RETURNS)															
Alpha	-0.027	-0.011	-0.005	-0.004	-0.002	-0.002	-0.001	-0.001	0.000	0.001	0.001	0.003	0.004	0.011	0.018
p-value-alpha	0.325	0.022	0.004	0.023	0.151	0.414	0.282	0.657	0.903	0.693	0.476	0.377	0.251	0.366	0.001
t-stat	-0.997	-2.400	-3.053	-2.300	-1.451	-0.822	-1.079	-0.445	-0.122	0.396	0.716	0.886	1.166	0.912	3.466
p-value-t-boot	0.799	0.225	0.013	0.033	0.083	0.155	0.001	0.000	1.000	0.830	0.693	0.783	0.713	0.934	0.217
Panel B: Baseline bootstrap statistics under single factor model (NET RETURNS)															
Alpha	-0.029	-0.011	-0.005	-0.004	-0.002	-0.002	-0.001	-0.001	0.000	0.001	0.002	0.003	0.004	0.011	0.017
p-value-alpha	0.273	0.019	0.372	0.137	0.023	0.505	0.203	0.307	0.946	0.729	0.353	0.241	0.013	0.349	0.001
t-stat	-1.111	-2.464	-0.904	-1.513	-2.291	-0.672	-1.282	-1.024	-0.067	0.346	0.932	1.177	2.524	0.944	3.375
p-value-t-boot	0.759	0.242	0.847	0.312	0.001	0.368	0.000	0.000	1.000	0.854	0.481	0.565	0.064	0.927	0.251

The table shows baseline bootstrap statistics at selected percentiles of the distribution of the three-factor and single-factor model for all UK-equity net fund returns over the sample period of 2002 to 2017. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) bootstrap's t-statistic estimate than the actual fund's t-statistic estimate.

8.2.2. Wild adjusted bootstrap Results

Table 8.4 shows wild adjusted bootstrap results for both single and three-factor models using funds' net returns. In Panel A, the three-factor model shows that the actual ranked alphas shift significantly to the left compared to those observed using gross returns in table 8.2. The t-statistic of ranked alpha is statistically significant for funds located below the 10^{th} percentile except from the bottom fund. We also observed statistically significant performance in the top performing fund at the 1% level. The p-value of the bootstrap shows that the significant performance cannot be explain by random sampling variation around a true value of zero. For example, the p-value of the bootstrap for percentiles of the CDF between the 1^{st} and 20^{th} quartiles is below the 5% significance cut off point. Hence, fund's performance at these points of the performance distribution are attributable to poor stock selection skill. However, the top performing fund was able to beat its luck distribution at the 1% level of significance.

Panel B of Table 8.4 shows the ex-post performance based on net returns under the single-factor model with a wild adjusted bootstrapped p-value at selected percentiles. Although the alpha performance is slightly lower than those seen under the three-factor model, the conventional t-test shows lower/higher statistical significance in the left/right tail of the distribution. For example, the p-value of actual alpha is statistically significant at the 1 % level for the 1^{st}, 20^{th} and 95^{th} percentile, and the top performing fund. Similarly, the p-value of the bootstrap reveals that the significant performance cannot be explain by random sampling variation around a true value of zero. For example, the t-statistic of actual alphas for these percentiles exceed those of the bootstrap simulations.

In conclusion, the evidence from wild adjusted bootstrap regarding skills versus luck is mixed. Introducing 1.95% annual fees across all funds to account for operating and management cost has pushed up the t-statistics for the actual fund alpha in the left tail of the performance distribution. Therefore, inferior stock selection skills are largely documented across the left tail of the performance distribution, regardless of the chosen model or the bootstrap scheme. In contrast, the t-statistics in the right tail of the distribution were significantly lower to those observed under gross return. This makes it a lot harder for funds' managers to beat their luck distribution. However, using the wild-adjusted bootstrap, our results suggest that there are a few fund managers who possess genuine stock picking skills and that can add value for their investors net of the operating and management fees.

.

Table 8.4: Statistical Significance of UK Equity Fund Performance.

Quantile	Min	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	Max
Panel A: Wild adjusted bootstrap statistics under three-factor model (NET RETURNS)															
Alpha	-0.027	-0.011	-0.005	-0.004	-0.002	-0.002	-0.001	-0.001	0.000	0.001	0.001	0.003	0.004	0.011	0.018
p-value-alpha	0.159	0.002	0.002	0.022	0.128	0.411	0.276	0.642	0.894	0.683	0.498	0.374	0.278	0.201	0.000
t-stat	-1.438	-3.273	-3.348	-2.315	-1.540	-0.828	-1.092	-0.467	-0.133	0.409	0.680	0.892	1.102	1.294	4.643
p-value-t-boot	0.718	0.050	0.002	0.012	0.035	0.104	0.000	0.000	1.000	0.653	0.628	0.683	0.679	0.618	0.000
Panel B: Wild adjusted bootstrap statistics under single factor model (NET RETURNS)															
Alpha	-0.029	-0.011	-0.005	-0.004	-0.002	-0.002	-0.001	-0.001	0.000	0.001	0.002	0.003	0.004	0.011	0.017
p-value-alpha	0.108	0.001	0.378	0.140	0.014	0.504	0.188	0.305	0.933	0.738	0.352	0.259	0.017	0.181	0.000
t-stat	-1.645	-3.535	-0.894	-1.499	-2.474	-0.674	-1.326	-1.029	-0.085	0.335	0.933	1.134	2.413	1.355	4.509
p-value-t-boot	0.674	0.029	0.814	0.252	0.000	0.233	0.000	0.000	0.999	0.775	0.328	0.476	0.046	0.638	0.000

The table shows wild adjusted bootstrap statistics at selected percentiles of the distribution of the three-factor and single-factor model for all UK-equity net fund returns over the sample period of 2002 to 2017. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) bootstrap's t-statistic estimate than the actual fund's t-statistic estimate.

8.3. Skills VS Luck Performance and Investment Style

In this section we examine whether fund manager's stock selection skills are influenced by funds' investment styles. The benefits of such an analysis are two-fold; first, it helps reduce bias in performance measurement due to model misspecification whereby grouping funds by their investment style would account for homogeneous risk across each investment style category. This helps control for cross-sectional risk characteristics by creating a benchmark that quantifies performance more effectively than the market index. Second, it helps us to identify whether fund manager stock selection skills are concentrated in certain investment styles. The bootstrap schemes are applied to four investment style categories, namely the (SG) small-growth, (SV) small-value, (BG) big-growth, and (BV) big-value investment styles. Information on funds' investment styles were obtained from the highest factor exposure produced by RBSA regression and followed the procedure used in the calendar time portfolios (see chapter 7 for more details). Thus, in February 2005, funds with a minimum of 36 observations are included within one of the four investment style categories, then we reset the clock in February 2010Accordingly, our analysis contains information on fund investment styles over 120-months (from February 2005- February 2015). These consist of 56 small-growth funds, 187 big-growth funds, 51 big-value funds, and 70 big-value funds. From chapter 7, we observed that the characteristic-based reference portfolio model[13] has a theoretical rationale and is better statistically determined than

[13] Characteristic-based reference portfolio model is a single factor of expected return. Where four reference portfolios were considered. SL represent the returns on value weighted portfolio of small-growth stocks. SH represent the returns on value weighted portfolio of small-value stocks. BL represent the returns on value weighted portfolio of big-growth stocks. BL represent the returns on value weighted

either the market benchmark model or the three-factor model. Accordingly, the discussion to follow is based on a characteristic-based reference portfolio model.

8.3.1. Performance and Investment Style-Baseline Bootstrap Results (Gross-Returns)

Table 8.5. reports the baseline bootstrap simulation results based on the t-statistics of a characteristic-based reference portfolio model alpha at selected quantiles for the four investment style categories. An important point to note is that the number of funds in each quantile may differ across investment styles. For example, there are 56 funds with a small-growth investment style objective, hence the 1st percentile contains only one fund, with the 1st percentile corresponding to the bottom performing fund (Min). Panel A of Table 8.5 shows that the small-growth ranked alpha performance from the characteristic-based reference portfolio model ranges from -1.8% to 0.2% per month for the lowest and top quantiles, respectively. However, the parametric test shows that the performance is only significant for the bottom performing fund at the 1% level. The bootstrap p-value result indicates that the performance of the bottom fund (0.01) and funds ranked between the 30th and 50th percentiles cannot be explained by random variation around the true value of zero alpha. Therefore, small-growth fund managers are found to exhibit poor stock selection ability at a 5% level of significance. Funds ranked beyond the 50th percentile of the performance distribution have failed to beat their luck distribution at a conventional level of significance.

---

portfolio of big-value stocks. (Obtained from Exeter University, Centre for finance and Investment (Gregory et al. 2013)).

Panel B of Table 8.5. shows that almost 80% of small-value funds have achieved positive alpha when using the characteristic-based reference portfolio model. However, the t-tests show that the gross alphas are statistically no different from zero, except for the top performing fund (statistically significant at the 5% level). The bootstrap p-values reveal that at the extreme tails of the distribution, small-value funds' performance is attributable to chance. However, funds ranked between the 60th and 80th percentile of the performance distribution are able to beat their luck distribution at a conventional level of significance.

Panel C reveals that 70% of the big-growth fund managers have achieved negative gross alpha for their investors. The t-tests show that only the bottom performing funds (1st percentile) yield significant underperformance at the 5% level. The bootstrap p-values indicate that inferior skill is highly significant for funds ranked between the 20th and the 50th percentile of the performance distribution.

In the case of big-value funds in Panel D, 80% of big-value fund managers deliver positive gross alpha. The standard t-tests show that the abnormal performance is only significant for the top percentile. The bootstrap p-values indicate that the top performing fund and funds ranked further inside toward the median of the performance distribution; particularly the 60th and the 80th percentiles which are skilful at a 1% significance.

From the four panels, it is clear that fund managers' skills are not equal between the four investment styles. The performance distribution is skewed toward the left for growth-oriented funds, while the performance distribution is skewed toward positive alpha for value-oriented funds. At the extreme tails of the performance distribution, the p-value of the bootstrap shows that poor stock selection skill is concentrated among the bottom 1 % of small growth funds. In contrast, superior skill is documented among the top 1% of big-value fund managers. Furthermore, across all investment styles, the p-value of the bootstrap is below the 5% cut-off point for funds who are ranked close to the median of the performance distribution. As we explained earlier, the residual variance of fund regression tends to be close to zero around the centre of the performance distribution. Thus, one would expect the actual t-statistic to exceed the 1,000 simulations bootstrap t-statistics.

8.3.2. Performance and Investment Style- Wild-Adjusted Bootstrap Results (Gross-Returns)

Table 8.6. presents the results of wild adjusted bootstrap p-values of the ranked t-statistics for small-growth, small-value, big-growth, and big-value funds, respectively.

The conclusions as described under the baseline bootstrap for the small-growth, small-value, and big-growth funds are unaltered. Panel A and C present an almost mirror image of the managerial skill/luck performance found in table 8.5 for growth-oriented funds. It shows that the poor performance found around the median of the distribution is attributed to poor stock selection skills rather than bad luck.

Table 8.5: Baseline bootstrap results under characteristic-based reference portfolio model.

Panel	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
Panel A: Small-Growth	Alpha	-0.018	-0.008	-0.006	-0.004	-0.003	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.002	0.002
	p-value-alpha	0.003	0.164	0.183	0.224	0.366	0.471	0.661	0.775	0.933	0.993	0.863	0.630	0.332
	t-stat	-3.108	-1.408	-1.347	-1.229	-0.911	-0.726	-0.441	-0.287	-0.084	0.009	0.173	0.484	0.978
	p-value-t-boot	0.041	0.660	0.393	0.090	0.059	0.021	0.014	1.000	1.000	1.000	1.000	0.999	0.974
Panel B: Small-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.005	-0.002	-0.001	0.000	0.001	0.001	0.002	0.003	0.003	0.005	0.005	0.006	0.007
	p-value-alpha	0.081	0.294	0.697	0.981	0.821	0.646	0.518	0.418	0.368	0.150	0.211	0.092	0.033
	t-stat	-1.782	-1.058	-0.391	0.024	0.228	0.461	0.650	0.816	0.906	1.458	1.265	1.712	2.181
	p-value-t-boot	0.634	0.888	1.000	1.000	1.000	1.000	1.000	0.002	0.051	0.026	0.424	0.267	0.219
Panel C: Big-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.008	-0.006	-0.005	-0.004	-0.004	-0.003	-0.002	-0.001	-0.001	0.001	0.002	0.003	0.005
	p-value-alpha	0.038	0.079	0.153	0.164	0.382	0.454	0.450	0.749	0.876	0.920	0.736	0.271	0.260
	t-stat	-2.128	-1.789	-1.450	-1.409	-0.882	-0.753	-0.760	-0.322	-0.157	0.101	0.339	1.111	1.137
	p-value-t-boot	0.385	0.349	0.386	0.037	0.056	0.001	0.000	1.000	1.000	1.000	1.000	0.814	0.984
Panel D: Big-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.004	-0.002	-0.002	0.000	0.001	0.002	0.003	0.003	0.004	0.005	0.007	0.008	0.010
	p-value-alpha	0.311	0.468	0.658	0.916	0.604	0.546	0.450	0.170	0.405	0.081	0.072	0.050	0.002
	t-stat	-1.021	-0.731	-0.445	0.106	0.521	0.610	0.762	1.389	0.839	1.794	1.831	2.026	3.263
	p-value-t-boot	0.985	0.971	0.999	1.000	1.000	1.000	0.999	0.000	0.119	0.004	0.105	0.176	0.011

The table shows baseline bootstrap statistics at selected percentiles of the distribution for the characteristic-based reference portfolio model. Gross funds returns are categorized by investment style as indicated in each panel over the sample period of 2005 to 2015. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) t-statistic's bootstrap estimate than the actual fund t-statistic's estimate.

However, the results with respect to the significance of the poor stock skills are strengthened. For example, the 1st percentile of the small-growth ranked fund (bottom fund) shows an actual ex-post skewness-adjusted t-statistic of -3.27 with a bootstrap p-value of 0.01. This indicates that only 1% of the bootstrapped t-statistics from 1000 simulations across all funds were lower than -3.27 under the null hypothesis of zero alpha. Hence, we can reject the hypothesis that the performance of the bottom fund is purely because of bad luck at the 1% level. The application of skewness-adjusted t-statistic and kurtosis preserving wild bootstrap shows improved inferences at the extreme tails of the performance distribution. The improved residual variance of fund regression is reflected in an improved t-statistic using the wild-adjusted bootstrap.

Similarly, Panel B presents superior skills for funds managers who are ranked between the 6th and the 8th quartiles, significant at the 1% level. With respect to big-value funds in Panel D, the significance level of the bootstrap p-values appears to be below the 5% cut off point for funds ranked at the right tail of the performance distribution. Beyond the 50th percentile, we can strongly reject the hypothesis that the performance is due to random sampling variation around a true value of zero. Thus, big value fund manager's possess genuine stock picking ability and can certainly beat their benchmark and add value for their investors gross of the operating and management fees. The skewness-adjusted and kurtosis preserving wild bootstrap shows stronger evidence of skilful performance among big-value fund managers.

By comparing the results across the four panels, we can conclude that there is strong evidence of inferior stock picking skill among funds with a growth- oriented investment style. Thus, regardless of the size effect, around 30% of growth-oriented funds, those that are ranked close to the median of performance distribution exhibit poor performance that cannot be regarded as bad luck. In contrast, superior stock selection talent is found to be statistically significant among 30% to 50% of value-oriented funds.

8.4. Empirical Results: Net Returns

8.4.1. Performance and Investment Style-Baseline Bootstrap Results (Net-Returns)

Table 8.7 reports the baseline bootstrap simulation results based on the ranked t-statistics of alphas of characteristic-based reference portfolio, regressed at selected quantiles for the four investment style categories. The abnormal performance based on net returns differ from those of gross returns by the average level of fund management fees (1.95% per year, or 0.16% per month). The conventional t-test shows that the performance is statistically different from zero at the 1st percentile for small-growth and small-value, and at the 99th percentile for big-value funds, while big-growth funds exhibit a highly significant underperformance up to the 20th percentile of the performance distribution. These results are identical to those observed under gross returns, apart from the significance level for small value and big-growth funds. Introducing operating expenses and management fees to gross returns has impacted the significance level of small-value and big-growth funds by lowering their t-statistics. The p-values of the bootstraps are very similar to those of the gross returns. It suggests strong evidence of underperformance in growth-oriented funds. Specifically, the actual t-statistics exceed the t-statistics of the 1000 simulations baseline bootstrap between the 20th and 50th percentile of the performance distribution.

Table 8.6: Wild adjusted bootstrap results under characteristic-based reference portfolio model.

Panel A: Small-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.018	-0.008	-0.006	-0.004	-0.003	-0.003	-0.002	-0.001	0.000	0.000	0.001	0.002	0.002
	p-value-alpha	0.002	0.149	0.194	0.208	0.365	0.470	0.651	0.754	0.926	0.994	0.862	0.641	0.334
	t-stat	-3.273	-1.462	-1.313	-1.272	-0.913	-0.727	-0.455	-0.315	-0.093	-0.007	0.175	0.469	0.974
	p-value-t-boot	0.014	0.434	0.265	0.012	0.023	0.005	0.006	1.000	1.000	1.000	1.000	1.000	0.941
Panel B: Small-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.005	-0.002	-0.001	0.000	0.001	0.001	0.002	0.003	0.003	0.005	0.005	0.006	0.007
	p-value-alpha	0.078	0.301	0.696	0.981	0.814	0.626	0.521	0.444	0.372	0.163	0.219	0.099	0.039
	t-stat	-1.801	-1.043	-0.393	0.024	0.236	0.490	0.645	0.771	0.900	1.412	1.243	1.677	2.114
	p-value-t-boot	0.449	0.826	1.000	1.000	1.000	1.000	1.000	0.001	0.017	0.005	0.250	0.104	0.116
Panel C: Big-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.008	-0.006	-0.005	-0.004	-0.004	-0.003	-0.002	-0.001	-0.001	0.001	0.002	0.003	0.005
	p-value-alpha	0.050	0.098	0.143	0.162	0.374	0.484	0.455	0.758	0.874	0.923	0.732	0.271	0.273
	t-stat	-1.998	-1.681	-1.486	-1.415	-0.897	-0.704	-0.752	-0.309	-0.159	0.098	0.344	1.112	1.107
	p-value-t-boot	0.377	0.290	0.176	0.003	0.013	0.001	0.000	1.000	1.000	1.000	1.000	0.741	0.978
Panel D: Big-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.004	-0.002	-0.002	0.000	0.001	0.002	0.003	0.003	0.004	0.005	0.007	0.008	0.010
	p-value-alpha	0.306	0.456	0.656	0.917	0.611	0.568	0.456	0.171	0.405	0.110	0.079	0.072	0.014
	t-stat	-1.034	-0.750	-0.448	0.104	0.511	0.576	0.752	1.385	0.838	1.638	1.790	1.852	2.587
	p-value-t-boot	0.931	0.934	0.987	1.000	1.000	1.000	1.000	0.000	0.014	0.000	0.007	0.056	0.037

The table shows wild-adjusted bootstrap statistics at selected percentiles of the distribution for the characteristic-based reference portfolio model. Gross funds returns are categorized by investment style as indicated in each panel over the sample period of 2005 to 2015. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Row 5 reports the corresponding p-value of the t-test in row 4. Finally, Row 6 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) t-statistic's bootstrap estimate than the actual fund t-statistic's estimate

In contrast, the performance of value-oriented funds can be explained by random sampling variation, except for the big-value funds who are located at the 60^{th} percentile of the distribution. The implication of these result is that UK equity funds cannot pick stocks well enough to cover their operating costs and management fees.

8.4.2. Performance and Investment Style- Wild-Adjusted Bootstrap Results (Net-Return)

Table 8.8 presents the ex-post performance based on net returns under the characteristic-based reference portfolio model with wild-adjusted bootstrapped p-values at selected percentiles. The alpha performance based on net returns is lower than those of gross returns by the average level of fund operating expenses and management fees. The conventional t-test shows identical results to those observed under the baseline bootstrap. Thus, the skewness-adjusted t-statistic has no impact on the power to detect abnormal performance. However, the p-value of the bootstrap shows that the significance of the poor stock skills is strengthened. For example, performance of growth-oriented funds below the 50^{th} percentile cannot be explained by random sampling variation around the true value of zero alpha, except for the 5^{th} and 1^{st} percentile of small-growth and big-growth funds, respectively. Therefore, growth-oriented funds managers are found to exhibit poor stock selection ability at a 5% level of significance. In such a case, one would conclude that almost 50% of growth-oriented funds cannot pick stocks well enough to cover their operating costs and management fees.

Table 8.7: Baseline bootstrap results under characteristic-based reference portfolio model (Net Returns).

	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
Panel A: Small-Growth	Alpha	-0.020	-0.009	-0.008	-0.006	-0.005	-0.004	-0.003	-0.003	-0.002	-0.002	-0.001	0.000	0.001
	p-value-alpha	0.001	0.094	0.091	0.099	0.180	0.256	0.377	0.476	0.608	0.647	0.761	0.971	0.758
	t-stat	-3.388	-1.704	-1.716	-1.678	-1.359	-1.148	-0.889	-0.718	-0.516	-0.460	-0.305	0.037	0.310
	p-value-t-boot	0.034	0.372	0.123	0.009	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel B: Small-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.007	-0.004	-0.003	-0.002	-0.001	0.000	0.001	0.001	0.002	0.003	0.004	0.004	0.006
	p-value-alpha	0.022	0.078	0.400	0.748	0.738	0.963	0.855	0.747	0.643	0.353	0.389	0.220	0.094
	t-stat	-2.373	-1.791	-0.848	-0.322	-0.335	-0.046	0.184	0.325	0.466	0.936	0.868	1.239	1.703
	p-value-t-boot	0.279	0.304	0.914	0.992	0.844	0.894	0.829	0.268	0.516	0.295	0.878	0.708	0.528
Panel C: Big-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.010	-0.008	-0.007	-0.006	-0.005	-0.004	-0.004	-0.003	-0.002	-0.001	0.000	0.002	0.004
	p-value-alpha	0.014	0.026	0.059	0.055	0.206	0.242	0.193	0.491	0.571	0.891	0.991	0.573	0.429
	t-stat	-2.545	-2.283	-1.929	-1.956	-1.282	-1.183	-1.318	-0.694	-0.570	-0.138	-0.011	0.567	0.797
	p-value-t-boot	0.164	0.104	0.086	0.001	0.001	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel D: Big-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.006	-0.004	-0.003	-0.001	-0.001	0.000	0.001	0.002	0.002	0.003	0.005	0.006	0.008
	p-value-alpha	0.160	0.227	0.382	0.630	0.769	0.951	0.807	0.506	0.521	0.230	0.166	0.116	0.010
	t-stat	-1.424	-1.220	-0.881	-0.485	-0.295	0.062	0.245	0.670	0.646	1.221	1.403	1.608	2.707
	p-value-t-boot	0.827	0.725	0.845	0.912	0.899	0.973	0.940	0.039	0.303	0.151	0.349	0.392	0.104

The table shows baseline bootstrap statistics at selected percentiles of the distribution for the characteristic-based reference portfolio model. Net funds returns are categorized by investment style as indicated in each panel over the sample period of 2005 to 2015. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Finally, Row 5 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) t-statistic's bootstrap estimate than the actual fund t-statistic's estimate.

The p-value of the wild adjusted bootstrap shows that the insignificance rates remain the same for small-value funds compared to the baseline bootstrap. We cannot reject the hypothesis that the net returns performance of small-value funds is due to chance at the selected percentiles. In contrast, the superior skill performance is more pronounced in big-value funds. The t-statistics of the 60th and 80th percentile ranked funds do exceed performance that might be explained by good luck at a 5% significance level. Thus, big-value funds which are located at these percentiles of the distribution seem to offer added value to their investors and to generate superior performance net of expenses and operating costs. Finally, by comparing these results with the baseline bootstrap, we find that the baseline bootstrap tends to attribute performance to chance more often than the wild-adjusted bootstrap.

8.5. Conclusion

In this Chapter, we evaluate the role of skill and luck in the performance of UK equity funds over the period 2002 to 2017. We compare the cross-sectional distribution of the t-statistics of alpha estimates for gross and net funds' returns against the simulated luck distribution. As a test of robustness in findings, the t-statistics of fund's alpha is estimated using the single and three-factor model. We also employ two bootstrap procedures to capture the luck distribution; namely the baseline bootstrap and the wild-adjusted bootstrap. The main advantage of the wild-adjusted bootstrap is that the luck distribution is constructed in such a way that mimics the four moments of the true fund's returns distribution. Thus, our wild adjusted bootstrap provides improved inferences in distinguishing skill from luck in fund performance.

Table 8.8. Wild adjusted bootstrap results under characteristic-based reference portfolio model (Net Returns).

Panel A: Small-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.020	-0.009	-0.008	-0.006	-0.005	-0.004	-0.003	-0.003	-0.002	-0.002	-0.001	0.000	0.001
	p-value-alpha	0.001	0.081	0.101	0.086	0.179	0.255	0.363	0.447	0.599	0.631	0.763	0.979	0.759
	t-stat	-3.583	-1.777	-1.666	-1.748	-1.361	-1.151	-0.916	-0.766	-0.529	-0.483	-0.304	0.026	0.308
	p-value-t-boot	0.008	0.199	0.040	0.001	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel B: Small-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.007	-0.004	-0.003	-0.002	-0.001	0.000	0.001	0.001	0.002	0.003	0.004	0.004	0.006
	p-value-alpha	0.020	0.084	0.398	0.749	0.746	0.979	0.857	0.764	0.645	0.365	0.396	0.228	0.102
	t-stat	-2.404	-1.755	-0.851	-0.322	-0.326	-0.026	0.181	0.302	0.463	0.912	0.855	1.218	1.659
	p-value-t-boot	0.119	0.157	0.824	0.984	0.778	0.891	0.888	0.233	0.362	0.167	0.747	0.488	0.360
Panel C: Big-Growth	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.010	-0.008	-0.007	-0.006	-0.005	-0.004	-0.004	-0.003	-0.002	-0.001	0.000	0.002	0.004
	p-value-alpha	0.021	0.039	0.052	0.054	0.197	0.278	0.198	0.503	0.568	0.888	0.995	0.572	0.440
	t-stat	-2.365	-2.116	-1.988	-1.967	-1.308	-1.096	-1.301	-0.673	-0.574	-0.141	-0.007	0.568	0.778
	p-value-t-boot	0.190	0.058	0.005	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel D: Big-Value	Quantile	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
	Alpha	-0.006	-0.004	-0.003	-0.001	-0.001	0.000	0.001	0.002	0.002	0.003	0.005	0.006	0.008
	p-value-alpha	0.154	0.214	0.379	0.628	0.763	0.966	0.822	0.506	0.521	0.262	0.174	0.144	0.032
	t-stat	-1.444	-1.255	-0.887	-0.487	-0.302	0.042	0.226	0.669	0.646	1.137	1.376	1.491	2.233
	p-value-t-boot	0.688	0.548	0.718	0.848	0.818	0.980	0.972	0.004	0.110	0.034	0.114	0.245	0.097

The table shows wild-adjusted bootstrap statistics at selected percentiles of the distribution for the characteristic-based reference portfolio model. Net funds returns are categorized by investment style as indicated in each panel over the sample period of 2005 to 2015. In each panel, the second row reports the actual alpha measured per month and sorted from worst performing fund(min) to best performing fund (max). The third row reports the corresponding p-value of the alpha in row two. Row 4 presents the corresponding t-statistics of the actual alpha estimate sorted from lowest (min) to highest (max). Row 5 reports the corresponding p-value of the t-test in row 4. Finally, Row 6 reports the percent of the 1000 simulation runs in each percentile that produce lower/higher (depending on if it is in left/right tail of the distribution) t-statistic's bootstrap estimate than the actual fund t-statistic's estimate

Furthermore, we investigate whether superior manager skill differs across different investment styles. Hence, investors can identify whether skill is concentrated within a certain investment style. The empirical evidence of the wild adjusted bootstrap suggests that, under the three-factor model and on a gross-returns basis, genuine stock selection skill is significant among the 95% and higher percentile of UK-equity funds. Thus, around 18 funds managers were able to beat their luck distribution at the 5% significance level. While the bootstrap results from the single factor model reports genuine stock picking ability in the 99^{th} percentile and top performing fund (around 4 funds). Furthermore, funds who are ranked slightly close to the median of the performance distribution (between 60^{th} and the 80^{th} percentile) have positive true alphas that cannot be explained by random sampling variation or chance. However, the residual variance (unexplained error) of fund regression tends to be close to zero around the centre of the performance distribution. Therefore, the 1000 simulations of the bootstrap coefficient estimates are more likely to have low sampling variation and produce a low t-statistic. Hence, the bootstrap methodology is most reliable at the extreme tails of the performance distribution when the variance of fund regression residuals is larger. Our conclusion is that, regardless of the expected gross returns model, there is no evidence of inferior stock picking skill among UK equity funds managers, while few fund managers have been able to beat their luck distribution. This result confirms what we would suspect, as we do not expect funds managers to deliberately make bad stock selections and underperform their benchmarks. It is also consistent with previously findings regarding the UK mutual fund market such as the work of Cuthbertson et al. (2008), and Blake et al. (2014).

When we examine ex-post performance once adjustments have been made to account for operating and management costs, the wild adjusted bootstrap shows that inferior stock selection skill does exist at the left tail of the performance distribution. The three-factor t-statistics of alpha estimates, for funds located between the 1^{st} and 20^{th} percentiles, exceeds the 1000 simulations in each selected percentile for more than 95% of simulation runs. Meanwhile, using the single factor model, the p-value of the wild-adjusted bootstrap is statistically significant at the 1^{st} and 20^{th} percentile. Thus, we conclude that around 10 to 20 percent of UK-equity funds managers do not have enough skill to cover their operating expenses and management fees. We also find that net returns set an obstacle for funds managers with positive alpha, in fact only the top performing fund is able to beat its luck distribution. Thus, any selectivity skills that funds managers might have are wiped out by operating and management fees, and their positive performance is simply due to chance. This result is inconsistence with the competitive model proposition of Berk and Green (2004), whereby rarely any skill is found in UK equity funds managers that is sufficient to cover their cost. However, our finding is largely consistent with Fama & French (2010), who report a lack of skills when fund's net returns are used. Nonetheless, there is the problem of drawing inferences from net returns. Our study employs a flat rate of operating and management fees across all funds, while in practice the amount of fees varies depending on funds' characteristics such as investor profile, strategy, size, and past performance.

By comparing the results across the two bootstrap approaches, we find that the baseline bootstrap uncovers much less good/bad skill in performance, particularly at the extreme tails of the performance distribution. However, we strongly favour the wild adjusted bootstrap as it mimics the properties of the parent distribution and provides improved inferences in distinguishing skill from luck in fund performance.

Finally, we investigate whether superior manager skill is concentrated in certain investment style, using a gross return t-statistic of alpha. Our results conclude that genuine stock picking talent is pronounced among value-oriented funds, whilst the bottom performing small-growth fund's performance is found to be worse than that which can be simply attributed to bad luck. Using a net return t-statistic of alpha, the result shows that funds managers do not possess enough skill to produce benchmark-adjusted net returns. Regardless of the bootstrap schemes, none of the investment styles are able to generate skilful performance in excess of operating expenses and management fees, except for the 80th percentile of the big-value investment style. Our evidence also suggests that the significant underperformance shown in growth-oriented funds is due to managers' inability to pick stocks well enough to cover their operating expenses and management fees. A similar conclusion was reached by Cuthbertson et al. (2008). Their findings tend to offer stronger support to a big-value investment style. However, our result is inconsistent with Chen et al. (2000), Barras et al. (2005) and Kosowski et al. (2006), where talent is prevalent in growth-oriented funds. It is worth noting that the t-statistics of alpha for growth-oriented funds are shifted to the left, almost 80% of our growth-oriented funds have a negative t-statistic of fund alpha. Fama & French (2010) show that in such a situation the true t-statistic of alpha estimate of skilled managers is most likely to be pushed down by poorly skilled managers who are extremely lucky. Therefore the performance of skilled managers is obscured by the performance of poorly skilled managers who are fortunate.

The findings of this chapter have two implications for investors in the UK equity funds. First, most UK equity funds' managers do not possess the superior skills that would allow them to add value for their investors after covering their operating costs and management fees. This raises doubt whether the cost of active fund managers is justifiable compared to passive funds. Second, investors who pursue a top ranked managed fund in their investment allocation might be worse off relative to passive investment strategies. Ranking funds according to their performance gives little information about the fund manager's stock picking talent or the fund's future performance.

Chapter 9

Conclusion

9.0. Introduction

The investment management industry has an important role in channelling savings to the capital market and is therefore a key source of funding for the UK economy. There are over £8.5trn of assets under management in the UK, of which £3.6trn of these assets are managed for overseas investors, with over half of this (2.1trn) coming from the EU. This makes the UK one of the most important centres for investment management, second only to the US in terms of size (Theia, 2020). To maintain global competitiveness and reputation, investment managers do not only have to deliver good performance to their investors but must also develop and produce products desired by investors (i.e., setting up a responsible and sustainable investment fund). From the investor's point of view, fund performance is very important to help making decisions about whom they ask to manage their money. Furthermore, investors are interested in how this performance is delivered, for example, whether the source of performance is the result of excessive risk taking or due to skill or luck. Although, many private firms and financial institutions provide rankings (ratings) of mutual funds' performance, selecting a mutual fund to best accommodate investor financial need is complex and requires expertise. The need for guidance in making this important choice has led to an increased demand for research studies that focus at evaluating the performance of mutual funds. In response to this need, our main objective is to comprehensively explore the performance of UK-equity unit trusts. Specifically, this thesis aims to analyse the ability of fund managers to select securities and add value above a set of style-adjusted benchmarks. We then turn to examine whether investors are able to exploit an ex-ante investment style strategy in the context of the event-time and the calendar time framework. Finally, we deal with the issue of data mining to assess whether fund managers abnormal performance is attributed to luck or skill.

In summary, this thesis offers a coherent, end-to-end picture of the unit trust performance measurement of the UK equity unit trust industry using a database of 352 trusts over the period Jan 2002 - December 2017. Our research makes a number of contributions to three areas of the existing literature on UK unit trust performance. First, we augment the commonly used factor models by using a style-adjusted benchmark that quantifies the performance more effectively than the general market index. While most previous work on mutual fund performance has limited its attention to factor models, we utilize style-adjusted model. Our approach not only improve performance measurement but also enable us to explore fund investment styles over an economic cycle. An important implication of our findings is that not using a style-adjusted model to evaluate funds can lead to an erroneous assessment of fund performance. Therefore, we advocate the use of style-adjusted benchmarks as a standard practice in mutual fund performance measurement.

Second, our discussion surrounding investors' ability to exploit a successful ex-ante investment style strategy is quite novel to fund performance literature and has an appealing feature for fund investors. We demonstrate that portfolio rebalancing inherent in standard performance measurement is misleading and doesn't capture investors true returns from a buy-and-hold strategy. We suggest evaluating fund performance in a way that is consistent with common investors holding period horizon, that is by style-adjusting the excess return of the fund over an investment horizon of a one- to five-year period. Our empirical evidence suggests that the results in current studies may be misstating unit trusts performance and investors' absolute returns. This is because they neglect that fund managers do not rebalance their portfolio as frequently as their benchmark does.

Third, we directly address the issue of skill versus luck in fund performance measurement. In our reappraisal of Kosowski et al. (2006) analysis, we replace the baseline bootstrap with the wild-adjusted bootstrap to determine the empirical distribution of idiosyncratic risk. We argue that the wild-adjusted bootstrap provides improved inferences in the evaluation of mutual fund performance, accounting for the non-normality and heteroscedasticity of individual mutual fund returns. We also demonstrate that skill and investment styles are mismatched among UK equity unit trust. In this sense, fund managers' stock selection ability is style varying. An important implication of our findings is that that the common practice of ranking funds according to their past performance (i.e., Morningstar's five–star rating system for mutual funds) gives little information about a fund manager's stock picking talent or a fund's future performance.

Our policy advice is twofold; first, our empirical evidence suggests that investment in actively managed fund in pursuit of abnormal performance, may well represent a misallocation of resources relative to passively managed fund. It is especially relevant given the state's pension deficit problem and government's encouragement of long-term saving using mutual funds. Second, current practice is for investors to select a fund based on past performance rather than managerial skill. We show that fund managers must have substantial stock picking skill, in order to compensate investors for the fees charged. In this sense, regulators should warn against trying to pick past winners and seek to provide independent information on fund's managers skill.

9.1. Research Questions and Objectives

Objective 1: To identify fund managers' stock selection behaviour in the context of return-based style analysis, then to evaluate the gross performance of both conventional and ethical funds.

In Chapter 5, we examined style preferences and the style-adjusted performance of UK equity unit trusts. We decomposed fund's returns into size and growth-value dimensions. Then we constructed 4 stylized portfolio and explored whether funds' performance differs across styles. Our result indicates that the bulk of conventional unit trusts have a tendency to favour big-oriented stocks. One possible explanation is that fund managers are aware of the difficulty in achieving long-term abnormal performance in an efficient market. Hence, they simply track the market index despite claiming otherwise. In term of performance, generally, the results reveal that on average UK-equity funds neither underperform or overperform their designated style benchmark. This finding is consistent with previous studies, such as Chan et al. (2002), Dimson et al. (2003) and Cuthbertson (2008), who report similar results after controlling for size and growth-value factor exposure. However, under continuous changing style portfolios formation, we document some evidence where investors can enhance style adjusted performance by investing in funds tilted toward small-value stocks. Nonetheless, performance is most likely to deteriorate with the introduction of transaction costs and management fees. The above results imply that using a style-adjusted benchmark provides a better explanation of the cross-sectional variation in UK-equity funds. The style-adjusted benchmark offers lower standard deviations and a higher adjusted R-squared than the market index (FTSE100).

In relation to ethical fund performance, under a continuous changing style approach, growth-oriented ethical funds have earned lower returns than their style benchmark indices. Furthermore, when performance is compared with conventional funds, ethical fund managers did worse than conventional fund managers on a style-adjusted basis. This result is consistent with previous empirical work on ethical funds in the UK market (Luther et al., 1992; Mallin et al., 1994; Gregory et al., 1997). However, our analysis shows that the disappointing performance cannot be blamed on ethical funds exposure to 'small firm' risk. With regards to the dominant style approach, the results tentatively show no support for ethical funds' underperformance compared to their conventional peers.

Objective 2: Test investors' ability to exploit an ex-ante investment style strategy to generate abnormal returns and provide empirical evidence on ethical fund investors' experience.

This task is accomplished in Chapter 6, which presents the results of the BHAR and CAR based on an investment strategy that could be achieved by systematically buying units in funds with a specific investment style objective over an investment horizon of a one- to five-year period. These tests are more appropriate for measuring investor's terminal wealth than the calendar time portfolio, and a wild-adjusted bootstrap is used to estimate the significance levels of the abnormal performance. The results show that, using a reference portfolio, value-oriented funds deliver positive abnormal returns for an investment horizon beyond 36 months. These results conform closely with the existing literature (Quigley and Sinquefield, 2000; Brookfield et al., 2013). In contrast, using the market index (FTSE 100), the results show no significant differences in returns on the short-and medium-term horizon. It is worth noting that, under an active investment strategy, investors are subject to higher charges than under a passive investment strategy. For example, the cost of active funds averages 1.95% per year compared to 0.84% for passive funds (Bryant and Taylor, 2012).

Chapter 6 also shows that in the event time, stylized ethical portfolios underperformed significantly, specifically when performance is measured relative to style-adjusted benchmarks. However, the underperformance disappears when performance is compared to the FTSE4GOOD index. More importantly, when performance is compared to a conventional funds' reference portfolio, we find no evidence of a statistically significant difference in returns between ethical and conventional fund returns. One exception is found in the small-value ethical funds, where the underperformance is noted at the long-term horizon. Specifically, the difference in returns ranges from -10% for the 3-year holding period to -26% for holding period of 5-year. These results are partially consistent with other reported studies in the UK ethical funds market. For example, Gregory and Whittaker (2007) and Bauer et al. (2005) have found no statistically significant differences in the performance of most ethical funds when compared to their conventional counterparts.

Objective 3: To control for cross correlations of abnormal returns and to assess whether performance is sensitive to the choice of empirical method and investment horizon.

The assumption of independence of abnormal returns within the sample may be too strong. The abnormal performances are likely to be non-random in the sample funds because of economic conditions and industry clustering. For example, fund managers might involve themselves in window dressing and style rotation activities to improve ex-post performance. Thus, if abnormal return is fund specific then the event-time portfolio return is not affected. However, if abnormal return is cross correlated, then inferences would be biased. The calendar time portfolio is suggested to correct for cross-sectional correlations, and to eliminate the bad model problem Fama (1998). Chapter 7 presents the results of the stylized calendar-time portfolios over a 5-year holding period, under the single and three-factor model. The results show no evidence of abnormal performance when funds are sorted on a style-adjusted basis. The results seem to be supportive of those in Lyon et al. (1999), Jegadeesh (2000), and Loughran and Ritter (2000), who argue against using the CTAR method as it has low power to capture abnormal returns and biased toward finding results consistent with market efficiency.

Finally, when comparing the performance of ethical funds relative to their conventional peers, the results show that the intercepts are negative and significantly different from zero as funds approach the value-orientation of the style spectrum. These results are partially corroborated with previous research on ethical fund performance (Gregory et al., 1997; Bauer et al., 2005; and Gregory et al., 2007). Furthermore, the ethical fund performance is expected to be worse when management fees are considered. For example, Bauer et al. (2005) and Bazo et al. (2008) show that ethical funds had significantly higher fees than conventional funds with similar characteristics. Overall, the results are in favour of the single factor characteristic-based model with GLS robust standard errors approach.

Objective 4: To separate out luck form skill (gross/net) performance and to identify whether fund managers stock selection skill is influenced by fund's investment style.

Up to Chapter 7, fund manager skill was discussed in the context of performance relative to a familiar benchmark model or relative to their peers. However, Kosowski et al. (2006) and Fama and French (2010) argue that this approach gives little information when it comes fund managers' skill and future performance. In order to avoid data snooping, two bootstrap procedures are employed, namely, the 'baseline', and the 'skewness-adjusted and kurtosis preserving wild' bootstraps. Using both the single and multi-factor model for net and growth returns, we have considered the role of skill and luck in the performance of UK equity funds.

Chapter 8 shows that the performance of the worst managers is solely due to bad luck rather than poor skill, while only a handful of funds display stock picking skill in the right tail of the distribution. At the extreme tail of the performance distribution, the three-factor model overstates the proportion of funds whose abnormal performance can be attributed to skill when compared against the single-factor model. However, once an allowance is made for management fees, inferior stock selection skills are largely documented across the left tail of the performance distribution. These results are consistent with those previously found in the UK mutual fund market such as, Cuthbertson et al. (2008) and Blake et al. (2014).

Chapter 8 also shows that fund manager stock selection skills are influenced by the fund's investment style. For example, the worst performing small-growth fund managers are not unlucky but rather unskilled, while a small group of skilled managers are documented in big-value funds. However, when net returns are considered, none of the UK- equity funds' managers possess sufficient skill to allow them to cover their operating costs and management fees. We therefore conclude that the selection of style-adjusted benchmark should be a vital consideration in the assessment of fund manager skill.

By comparing the results across the two bootstrap approaches, we find that the baseline bootstrap uncovers much less good/bad skill in performance, particularly at the extreme tails of the performance distribution. However, we strongly favour the wild adjusted bootstrap as it mimics the properties of the parent distribution and provides improved inferences in distinguishing skill from luck in fund performance.

9.2. Limitations and Recommendation for Future Research

As with most studies of the UK unit trusts performance, the main limitation of this study is dictated by data availability. Data collection has been a troublesome procedure and constrained by the lack of fund information. For example, commercial databases drop fund's information once the fund has exited the market, creating a survivorship bias in the results. Consequently, data was collected manually from several sources. However, information on fund size, fund flows, and expenses information were unattainable, especially for those funds which no longer exist.

Berk and Green (2004) point out that as the size of the active mutual fund increases, a fund's ability to outperform the passive benchmark declines. Hence mutual funds exhibit significant diseconomies of scale in performance. Droms and Walker (1996), Wermers (2000) and Garyn-tal (2015) show that there is a significant relationship between mutual funds abnormal performance and their expenses ratio. Bauer et al. (2005) and Bazo et al. (2008) indicate that ethical funds have significantly higher fees than conventional funds. Given that these fund characteristics are an integral part of the fund's performance, they cannot be looked at in isolation. Because of this further research comparing the impact of fund size, fund flows, and the expenses ratio to fund performance will be needed. This thesis can be further extended to include a longer sample period as some literature argue that there is a relationship between the length of the performance record and the power of the test for assessing fund management skills. For example, Blake and Timmermann (2001) show that it takes approximately 8 years of performance data for a test of a fund manager's skill to have 50% power and 22 years of data for the test to have 90% power.

With regards to the return-based style analysis, the factor loadings are computed over a 36-month rolling window. Although this window is sufficient to provide an accurate estimate of the factor loadings, it gives equal weighting to the 36-month returns. A potential avenue for further research is to apply a lower frequency of data (i.e., weekly) and shorter window (i.e., 52 weeks). This would provide an accurate analysis of investment style and support the identification of a mean-variance efficient benchmark.

With regards to the event time performance, although the benchmark reference portfolio is a plausible and investable opportunity, it is not an easily replicable strategy. Thus, it would be of interest to test against an ETF (Exchange traded fund) that matches the style/risk of an event portfolio. With regards to the calendar time approach, we measure performance using a regression-based framework. The expected return is estimated using the characteristic-based reference portfolio, the market benchmark, and the Fama-French three-factor model. Therefore, the validity of the results is dependent upon the reliability of risk-adjusted benchmarks in explaining the cross-section of expected returns. A potential avenue for further research would be to test whether alternative asset pricing models can provide a better explanation of the cross-sectional variation in UK-equity funds.

Finally, when separating out skill from luck in performance, both baseline and wild-adjusted bootstrap ignore the systematic relationship between a fund's returns and the factor benchmarks. Furthermore, by randomly sampling across individual fund residuals, we lose any effects of autocorrelation in returns. To our knowledge, the first problem cannot be addressed unless all funds exist at the same time (then we can jointly sample fund residuals and explanatory returns). However, the second problem can be addressed using block or stationary bootstrap. Thus, a further suggestion for future research is to investigate manager skill using a block or stationary bootstrap for a robust result with respect to this alternative procedure.

References

Al-Horani, A., Pope, F., and Stark, A. 2003. Research and Development Activity and Expected Returns in the United Kingdom. European Finance Review. Vol. 7, pp. 27-46.

Ali, A., Lee-Seok, H., and Mark T. 2003,b. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, 69, pp. 355-373.

Almazan, A., Brown, K., Carlson, M., and Chapman, D. 2004. Why constrain your mutual fund manager?. *Journal of Financial Economics*, 73 (2), pp. 289-321.

Annaert, J., and Van Campenhout, G. 2002. Style Breaks in Return-Based Style Analysis. *University of Antwerp*, Department of Applied Economics. Available from: https://www.researchgate.net/publication/4983020_Style_breaks_in_return-based_style_analysis/link/02e7e516c0b69d87a0000000/download.[Accessed 05/06/2018].

Angelidis, T., Giamouridis, D., and Tessaromatis, N. 2013. Revisiting mutual fund performance evaluation. Journal of Banking and Finance. Vol. 37(5), pp. 1759–1776.

Bassett, G., and Chen, H. 2001. Portfolio style: Return-based attribution using quantile regression. *Empirical Economics*, 26, pp. 293-305.

Bacquero, J., Horst, J., and Verbeek, M. 2002. Survival, Look-Ahead Bias and the Persistence in Hedge Fund Performance. *Erasmus research institute of management report*. Accessed from: file:///Users/Aboud/Downloads/erimrs20021119130344.pdf.

Baker, M., Litov, L., Wachter, J., and Wurgler, J. 2010. Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements. *Journal of Financial and Quantitative Analysis,*45(5), pp. 1111-1131.

Bangassa, K., Su,C., Joseph, N. 2012. Selective and timing performance of UK investment trust. *Journal of International Financial Markets, Institutions & Money*, 22, pp. 1149-1175.

Banz, R. 1981. The relationship between returns and market value of common stock. *Journal of Financial Economics,* 9, pp. 3-18.

Barber, B., Lyon, J. 1997. Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics. *Journal of Financial Economics,* 43, pp. 341-372.

Barberis, C., Shleifer, A., and Vishny, W. 1998. A model of investor sentiment. *Journal of Financial Economics,* 49, pp. 309-343.

Barber, M., and Lyon, D. 1997. Detecting Long-Run Abnormal Stock Performance: The Empirical Power and Specification of Test Statistics. *Journal of Financial Economics*, 43, pp. 341-372.

Basu, J. 1983. The relationship between earnings, Yield, Market value and Returns. *Journal of Financial Economics*, 12, pp.129-156.

Baur, R., Kees, K., and Roger, O. 2003. International Evidence on Ethical Mutual Fund Performance and Investment Style. *Journal of Banking & Finance,* 29, pp. 1751–1767.

Bauer, R., Koedijk, K., and Otten, R. 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7), pp.1751-1767.

Becker, C., Ferson, W., Myers, D. and Schill, M. (1999). Conditional market timing with benchmark investors, Journal of Financial Economics, 52, pp 119-148.

Berk, J., and Green, R. 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy,* 112, pp. 1269-1295.

Ben Dor, A., R. Jagannathan. (2003). Understanding Mutual Fund and Hedge Fund Styles Using Return-Based Style Analysis. *Journal of Investment Management*, 1(1), pp. 94-134.

Betker, B., and Sheehan, J. 2013. A comparison of single factor and multiple factor alphas used in measuring mutual fund performance. *Financial Services Review*, 22, pp. 349-365.

Blake, D., Caulfield, T., and Ioannidis, C. 2014. New Evidence on Mutual Fund Performance: A Comparison of Alternative Bootstrap Methods. *Journal of Financial and Quantitative Analysis.*

Blake, D, Timmermann, A.1998. Mutual fund performance: Evidence from the

UK. European Finance Review. 2. pp. 57-77.

Blake, D and Timmermann, a. 2001. Performance Benchmarks for Institutional Investors; Measuring, Monitoring and Modifying Investment Behaviour. Discussion Paper PI-0106, The Pensions Institute. Available from: file:///C:/Users/User/Downloads/Performance_Benchmarks_for_Institutional_Investors.pdf. Accessed on 07/06/2020

Brav, A. 2000. Inference in Long-Horizon Event Studies: A Bayesian Approach with Application to Initial Public Offerings. *Journal of Finance*, 55, (5).

Brav, A., Geczy, C., and Gompers, P. 2000 Is the abnormal return following equity issuances anomalous? Journal of Financial Economics. Vol. 56, pp. 209-249.

Brookfield, D., Su, C., and Bangassa, K. 2015. Investment style positioning of UK unit trusts. *The European Journal of Finance*, 21, (10–11), pp. 946-970.

Brown, J, Goetzmann, W, and Ibbotson, G. 1992. Survivorship bias in performance

studies. The Review of Financial Studies. 5 (4). pp. 553:580.

Brown, J., and Goetzmann, W. 1997. Mutual Fund Styles. *Journal of Financial Economics*. 43, pp. 373-399.

Brown, J., Goetzmann, W., and Ibbotson, G. 1992. Survivorship bias in performance

studies. The Review of Financial Studies, 5 (4), pp. 553-580.

Brown, S., and Warner, J. 1980. Measuring security price performance. *Journal of Financial Economics*, 8, pp. 205-258.

Brown, S., and Warner, J.1985. Using daily stock returns: the case of event studies. *Journal of Financial Economics*, 14, pp.3-31.

Brown, S., Goetzmann, W. 1995. Performance Persistence. *The journal of Finance*, 50(2), pp. 679-698.

Brown, S., Goetzmann, W, Ibbotson, R, and Ross, S. 1992. Survivorship bias in

performance studies. *Review of Financial Studies*,5, pp. 553-580.

Brown, G., P. Draper, and E. McKenzie. 1997. Consistency of UK pension fund performance. Journal of Business Finance and Accounting, pp.155–78.

Bryant, C., and Taylor, G. (2012). Fund Management Charges, Investment Costs and Performance. IMA Statistics Series Paper 3. Available online from: https://www.theia.org/sites/default/files/2019-05/20120504-imastatisticsseriespaper3.pdf. Accessed on the [01/06/2018].

Byrne, A., Fletcher, J., and Ntozi, P. 2006. An exploration of the conditional timing performance of UK unit trusts, *Journal of Business Finance and Accounting*, 33 (5) & (6), pp. 816-838.

Campbell, C., and Wasley, C. 1993. Measuring security price performance using daily NASDAQ returns, *Journal of Financial Economics*,33, pp. 73-92.

Carhart, M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1), pp. 57-82.

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

Chan, L., Chen, H., and Lakonishok, J. 2002. On Mutual Fund Investment Styles. *Review of Financial Studies*,15 (5), pp. 1407-1437.

Chevalier, J., & Ellison, G. 1999. Are some mutual fund managers better than others? Cross-sectional patterns in behaviour and performance, *Journal of Finance*, 54(3), pp. 875-899.

Chan, L., Chen, H., and Lakonishok, J. (2002). On Mutual Fund Investment Styles. *Review of Financial Studies*, 15(5), pp. 1407–1437.

Chen, L., Adams, A., and Taffler, R. 2011. What style-timing skills do mutual fund “stars” possess?. *Journal of Empirical Finance*, 21, pp.156–173

Chen, Z., and Knez, P. 1996. Portfolio performance measurement: theory and applications, *Review of Financial Studies*, 9 (2), pp. 511-55.

Chen, H., Jegadeesh, N., and Wermers, R. 2000. The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers. T*he Journal of Financial and Quantitative Analysis*, 35 (3), pp. 343- 368.

Chen, J., Hong, H., Huang, M. and kubik, J. 2002. Does Fund Size Erode Performance? Liquidity, Organizational Diseconomies and Active Money Management. *Syracuse University - Department of Economics.*

Chevalier, J., and Ellison, G., 1997. Are some mutual funds managers better than others? Cross-sectional patterns in behavior and performance. *Journal of Finance*, 54, pp.562-571.

Christopherson, J., Ferson, and Glassman, D. (1998). Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance. *Review of Financial Studies*, 11(1), pp.111-142.

Clash, J. (1998). The wallflower strategy. Forbes, pp. 113-117.

Collins, D., and Dent, W. 1984. A Comparison of Alternative Testing Models Used in Capital Market Research, *Journal of Accounting Research*, 22, pp. 48-84.

Corrado, J., and Zivney, L. 1992. The specification and power of the sign test in event study hypothesis tests using daily stock returns. *Journal of Financial and Quantitative Analysis*, 27, pp. 465-478.

Corrado, J. 1989. A nonparametric test for abnormal security performance in event studies. *Journal of Financial Economics*, 23, pp. 385-395.

Cowan, R. 1992.Nonparametric event study tests. *Review of Quantitative Finance and Accounting*, 2, pp. 343-358.

Cuthbertson, K., Nitzsche, D., and O’Sullivan, N. 2010. The Market Timing Ability of UK Mutul Funds. *Journal of Business Finance and Accounting*, 37(1), pp. 270-289.

Cuthbertson, K., Nitzsche, D., O’sullivan,N. 2010. Mutual Fund Performance: Measurement and Evidence, *Journal of Financial Markets, Instruments and Institutions,*19(2), pp.95-187.

Cuthbertson, K., Nitzsche, D., O'sullivan, N. 2006. Mutual fund performance. Cass Business School, City University. Working paper. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=955807. [Accessed on 03 March, 2015].

Cuthbertson, K., Nitzsche, D., O'Sullivan, N. 2008. UK mutual fund performance: Skill or luck?. *Journal of Empirical Finance*, 15, pp. 613-634.

Cuthbertson, K., Nitzsche, D., O'Sullivan, N. (2012). False Discoveries in UK Mutual Fund Performance. European Financial Management, Vol. 18, No. 3, 2012, 444–463.

Cuthbertson, K., Nitzsche, D., O'Sullivan, N. (2016). A review of behavioural and management effects in mutual fund performance. International Review of Financial Analysis. Vol 44. pp. 162–176.

Daniel, G., Sornette, D., Woehrmann, P. 2009. Look- Ahead Benchmark Biase in Portfolio Performance Evaluation. *The journal of Portfolio Management,* 36 (1), pp. 121-130.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance,* 52(3), pp. 1035-1058.

Davidson, R., Duclos, Y. 2000. Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econometrica*, 68, pp.1435-1464.

Davidson, J., Monticini, A., and Peel, D. 2007. Implementing the Wild Bootstrap Using a Two-Point Distribution. *Economics Letters*, 96, pp. 309–15.

Davis, J. 2001. Mutual fund performance and manager style. *Financial Analysts Journal*, 57, pp. 19-27.

Dimson, E., Nagel, S., and Quigley, G. 2003. Capturing the Value premium in the UK 1955-2001. *Financial Analysts Journal*, 59, pp. 35-46.

Eckbo, E., Masulis, W., & Norli, O. 2000. Seasoned public offerings: Resolution of the "new issues puzzle. Journal of Financial Economics. Vol. 56, pp. 251–291.

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

EIRIS. 2021. EIRIS Foundation Responsible Investment in Charity Pooled Funds. Available from: https://eirisfoundation.org/wp-content/uploads/2021/02/Charity-Pooled-Funds-Report.pdf. [Accessed on 02 March, 2022].

Fama, E. F. (1965), Random Walks in Stock Market Prices, *Financial Analysts Journal* 21, PP.55–59.

Fama, E, 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance,*25 (2), pp.383-417.

Fama, E. (1976) Foundation of Finance. The Journal of Finance. New York: Basic Books. Available from: https://altexploit.files.wordpress.com/2017/09/eugene-f-fama-foundations-of-finance_-portfolio-decisions-and-securities-prices-1976-06-1976.pdf. [Accessed on 03 March, 2017]

Fama, E., Fisher, L., Jensen, M., Roll, R. 1969. The adjustment of stock prices to new information. *International Economic Review*, 10, pp. 1-21.

Fama, E, French, K. 1993. Common Risk Factor in the Return on Stocks and Bonds. *Journal of Financial and Economics*, 33, pp. 3-56.

Fama, E, French, K. 2010. Luck versus Skill in the Cross-Section of Mutual Fund Returns. *Journal of finance*, LXV (5), pp. 1915- 1947.

Fama, E. (1972). Components of Investment Performance. *Journal of Finance*, 27 (3), pp.551-567.

Fama, E. 1998. Market efficiency, long-run returns, and behavioural finance. *Journal of Financial Economics*, 49(3), pp. 283-306.

Fama, E. F., and K. R. French. 2010. "Luck versus Skill in the Cross-Section of Mutual Fund Returns. Journal of Finance, 65, 607–636.

Fama, E, and French, K. 2015a. A Five-factor Asset Pricing Model. Journal of Financial Economics. 116, pp.1–22.

Feng, X., Johansson, A. 2015. Can mutual funds pick stocks in China? Evidence from the IPO market. *Journal of Banking & finance,* 55, pp. 170-186.

FCA. (2019). *The Collective Investment Scheme Information Guide*. Available from:https://www.handbook.fca.org.uk/handbook/COLLG.pdf.[Accessed 05/09/2018].

Ferson, W., Schadt, R. (1996). Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance*, 51 (2), PP, 425-461.

Fletcher, J. 1997. An examination of UK unit trust performance within the arbitrage pricing framework. *Review of Quantitative Finance and Accounting*, 8, pp.91- 107.

Fletcher, J., and D. Forbes.2002. An exploration of the persistence of UK unit trust

FSA Occasional Paper, 9, pp.1-58.

Fletcher, J. 2015. Exploring the persistence of UK. equity closed-end fund performance. University of Strathclyde. Department of Accounting and Finance, https://strathprints.strath.ac.uk/52796/1/Fletcher_AIAPM_2015_Exploring_the_persistence_of_UK_equity_closed_end.pdf

Foran, Jason, and Niall O'Sullivan. 2017. Mutual Fund Skill in Timing Market Volatility and Liquidity. International Journal of Finance and Economics 22: 257–73

Gregory, A., Matatko, J., Luther, R., 1997. Ethical unit trust financial performance: Small company effects and fund size effects. *Journal of Business Finance & Accounting*. Vol. 24, pp. 705–725.

Gregory, A., and Whittaker, J. 2007. Performance and Performance Persistence of 'Ethical' Unit Trusts in the UK. *Journal of Business Finance & Accounting*, 34(7&8), pp. 1327–1344.

Gregory, A., Tharyan, R., & Christidis, A. 2013. Constructing and testing alternative versions of the Fama–French and Carhart models in the UK. *Journal of Business Finance & Accounting*, 40(1&2), pp.172-214.

Gergory, A., Guermat, C., and Al-Shawawreh, F. 2010. UK IPOs: Long Run Returns, Behavioural Timing and Pseudo Timing. *Journal of Business Finance & Accounting*, 37(5) & (6). pp. 612-647.

Gregory, A., Tharyan, R., & Christidis, A. (2013). Constructing and testing alternative versions of the Fama–French and Carhart models in the UK. *Journal of Business Finance & Accounting*, Vol.40(1&2), pp.172-214.

Goetzmann, W., Ingersoll, J., and Ivkovich, Z. 2000. Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis*,35, pp. 257 -290.

Green, R. 1986. Benchmark Portfolio Inefficiency and Deviations from the Security Market Line. *Journal of Finance,* 41, pp. 295-312.

Gregory, A., and Whittaker, J. 2007. Performance and Performance Persistence of 'Ethical' Unit Trusts in the UK. *Journal of Business Finance & Accounting*,34(7) & (8), pp.1327-1344.

Gregory, A., Matatko, J., and Luther, R. 1997. Ethical unit trust financial performance: Small company effects and fund size effects. *Journal of Business Finance and Accounting,* 24 (5), pp. 705-724.

Grinblatt, M., Titman, S. 1989. Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings. *The Journal of Business*, 62(3), pp. 393-416.

Grinblatt, M., Titman, S. 1992. The persistence of mutual fund performance. *Journal of Finance*, 47 (5), PP, 1977-1984.

Grossman, S., and Stiglitz, J. 1980. On the impossibility of informationally efficient markets. *The American Economic Review*, 70 (3), PP, 393-408.

Guerard, B. 1997. Is there a cost to being socially responsible in investing?. *The Journal of Investing*, pp.11-18.

Hamilton, S., Jo, H., and Statman, M. 1993. Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, 49 (6), pp. 62-66.

Henriksson, R., Merton, R. 1981. On Market Timing and Investment Performance II Statistical Procedures for Evaluating Forecasting Skills. *Journals of Business*, 54 (4), pp. 513-533.

Henriksson, R. 1984. Market Timing and Mutual Fund Performance: An Empirical Investigation. *Finance Working Paper No.137*. Available from: http://www.haas.berkeley.edu/groups/finance/WP/rpf137.pdf. [Accessed on 03 March, 2018].

Hong, H., Kacperczyk, M. 2005. The price of sin: the effect of social norms of markets. Working Paper Princeton University.

Hooks, J. 1996. The Effects of Loads and Expenses on Open End Mutual Fund Returns. *Journal of business research*, 36 (2), pp. 199-202.

Huang, R., A, D., & Poulot, W. 2020. A reappraisal of luck versus skill in the cross-section of mutual fund returns. *Journal of Economic Behavior & Organization*, 176, pp.166-187.

Hutton, P. 1997. Mori Customer Relation Research: Using Research to improve Quality and Service Provision, Paper at S M I Conference.

IMA Investment Management Association. (2012). Enhanced disclosure of fund charges and costs. Available online from: https://www.theia.org/sites/default/files/2019-06/20120920-enhanceddisclosureoffundchargesandcosts.pdf. Accessed on the [01/06/2018]

IMA. 2016. The investment Management association. Retrieved from: http://www.theinvestmentassociation.org/fund-statistics/. [Accessed on 03 March 2018]. Ikenberry, D., Lakonishok, J., Vermaelen, T. 1995. The Underreaction to Open Market Share Repurchases. *Journal of Financial Economics,* 39, pp. 181-208.

Ippolito, R. 1989. Efficiency With Costly Information: A Study of Mutual Fund Performance 1965-1984, *The Quarterly Journal of Economics*, 104, pp. 1-23.

Jaffe, J. 1974. Special information and insider trading. *Journal of Business*,47, pp. 411-428.

Javier, G. 2013. The persistence of European mutual fund performance. Research in International Business and Finance. Vol 28, pp. 45-67

Jegadeesh, N., and Titman, S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, pp. 65-91.

Jegadeesh, N., and Karceski, J. 2004. Long-run performance evaluation: Correlation and heteroskedasticity-consistent tests. working paper, Emory University.

Jensen, M. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance*, 23, pp. 389-416.

Jiang, W. (2003), A nonparametric test of market timing, *Journal of Empirical Finance*. 10 (4). pp. 399-425.

Jonathan, B., and Binsbergen, J. 2015. Active Managers Are Skilled. Stanford University Graduate School of Business Research Paper No. 15-37. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2616505.

Kosowski, R., Timmermann, A., White, H., Wermers, R. 2006. Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, LXI (6), pp.2551-2595.

Keswani, A. Stolin, D. 2008. Which money is smart? mutual fund buys and sells of individual and institutional investors. Journal of Finance. 63 (1), pp. 85:118.

Kothari, P., and Warner, J. 1997. Measuring long-horizon security price performance. *Journal of Financial Economics*, 43, pp.301-339.

Kothari, s., and Warner, J. 2006. Econometrics of Event Studies. *Tuck School Handbook of Business at Dartmouth*. Corporate Finance: Empirical Corporate Finance. Available from: http://www.bu.edu/econ/files/2011/01/KothariWarner2.pdf. [Accessed on 12 August 2016].

Kosowski, R, Timmermann, A, White, H, Wermers, R. 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. Journal of Finance. LXI (6), pp. 2551:2595.

Kreander, N., R. H. Gray, D. M. Power and C. D. Sinclair.2005. Evaluating the Performance of Ethical and Non-Ethical Funds: A Matched Pair Analysis. Journal of Business Finance & Accounting. Vol. 32, (7&8), pp. 1465–93.

Lakonishok, A., Shleifer, R., and Vishny, R. 1991.Window dressing by pension fund managers. *American Economic Review,* 81, pp.227-258.

Lakonishok, J., Shleifer, A., and Vishny, R. 1994. Contrarian investment, extrapolation, and risk, *Journal of Finance*,49, PP,1541-1578.

Liu, Y. 1988. Bootstrap Procedure Under Some Non-i.i.d. Models. Annals of Statistics, 16, pp. 1696–708.

Liu, W., and Strong, N. 2008. Biases in Decomposing Holding-Period Portfolio Returns. *Review of Financial Studies*, 21 (5), pp. 2243-2274.

Lisi, F., and Caprin, M. 2012. On the role of risk in the Morningstar rating for mutual funds. Quantitative finance, Vol 12(10), pp. 1477-1486.

Lo, A., and MacKinlay, C. 1990. When are Contrarian Profits Due to Stock Market Overreaction?. *Review of financial studies*, 3, pp. 175-205.

Loughran, T., Ritter, J. 2000. Uniformly Least Powerful Tests of Market Efficiency. *Journal of Financial Economics*, 55 (3), pp. 361:389.

Lobosco, A., and DiBartolomeo, D. 1997. Approximating the Confidence Intervals for Sharpe Style Weights. *Financial Analysts Journal*, 53, pp. 80–85

Loughran, T., and Ritter, J.2000. Uniformly least powerful tests of market efficiency. Journal of Financial Economics. Vol. 55: 361-389.

Luther, R., and Matatko, J. 1994. The Performance of Ethical Unit Trusts: Choosing an Appropriate Benchmark. *British Accounting Review*, 26, pp. 57–70.

Luther, R., Matatko, J., Corner, D., 1992. The investment performance of UK ethical unit trusts. *Accounting, Auditing and Accountability Journal*, 5 (4), pp. 57-70.

Lyon, J, and Barber, B. 1997. Detecting long-run abnormal stock returns: the empirical power and specification of test statistics. *Journal of Financial Economics,* 43, pp. 341-372.

Lyon, J, Barber, B, and Tsai, C. 1999. Improved methods of tests of long-horizon abnormal stock returns. *Journal of Finance, 54*, pp. 165-201.

Mallin, A., Saadouni, B., and Briston, R. 1995. The Financial Performance of Ethical Investment Funds. *Journal of Business Finance & Accounting*, 22(4), pp. 483–96.

Mateus, I., Cesario, a., Todorovic, N. 2016. UK equity mutual fund alphas make a comeback. International Review of Financial Analysis. Vol 44. pp. 98–110.

Mamaysky, H., Spiegel, M. and Zhang, H. (2007). Improved forecasting of mutual fund alphas and betas. Review of Finance, Vol, 11, pp. 359–400.

Mandelker, G. 1974. Risk and return: The case of merging firms. *Journal of Financial Economics*, 1, pp. 303-335.

Massa, M., and Patgiri, R. 2009. Incentives and mutual fund performance: higher performance or justhigher risk taking?. *Review of Financial Studies*. 22 (5). pp. 1777-1815.

Mitchell, M., Stafford, E. 2000. Managerial decisions and long term stock price performance. *Journal of Business,* 73 (3), pp. 287-329

Morningstar. 2018. Halifax UK Growth Fund B. Available from:

https://www.morningstar.co.uk/uk/funds/snapshot/snapshot.aspx?id=F0GBR05B97. [Accessed 05/09/2018].

Morningstar. 2018,a. Halifax UK Growth Fund B. Available from:

https://www.morningstar.co.uk/uk/funds/snapshot/snapshot.aspx?id=F0GBR05B97. [Accessed 05/09/2018].

Morningstar (2018,b). BMO Responsible UK Equity Fund. Available from: https://www.morningstar.co.uk/uk/funds/snapshot/snapshot.aspx?id=f0gbr04dhw [Accessed 05/09/2018].

Nain, A., Yao, T. 2013. Mutual Fund Skill and the Performance of Corporate Acquirers. *Journal of Financial Economics*, 110 (2), pp. 437-456.

Ornelas, J., Silva, A., Fernandes, J. (2012). Yes, the choice of performance measure does matter for ranking of US mutual funds. international Journal of finance and economics. Vol 17 pp. 61-72

Patell, J. 1976. Corporate Forecasts Of Earnings Per Share And Stock Price Behavior: Empirical Tests. *Journal of Accounting Research*, pp. 246-276.

Pastor, L, Stambaugh, R,. 2001. Liquidity Risk and Expected Stock Returns. *NBER Working Paper No. w8462*

Pettengill, G., Chang, G., and Hueng, J. (2014). Choosing between value and growth in mutual fund investing. *Financial Services Review*. Vol, 23 pp. 341-359.

Pettengill, G., Chan, G., and Hueng, J.2014. Choosing between value and growth in mutual fund investing. *Financial Services Review*, 23, pp. 341-359.

Pastor, l., Stambaugh, R.F. and Taylor, L.A., 2015. Scale and skill in active management. Journal of Financial Economics, 116(1), pp. 23-45.

Quigley, G., and Sinquefield, R. 2000. Performance of UK equity unit trusts. *Journal of Asset Management*, 1, pp. 72-92.

Reilly, F., and Brown, K. 2015. Analysis of investments and management of portfolios. *Publisher: Andrew Ashwin*, 1st edition, PP,330-331.

Reinganum, R. 1981.Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values. *Journal of Financial Economics,* 9, pp. 19-46.

Rhodes, M. 2000. Past imperfect? The performance of UK equity managed funds. *Review of Financial Studies*, 18(2), pp. 569-597.

Ritter, J. 1991. The Long-run Performance of Initial Public Offerings. *Journal of Finance,* 46, pp. 3-27.

Roll, R. 1978. Ambiguity when performance is measured by the securities market line. *Journal of Finance*, 33 (4), pp. 1051-1069.

Ross, S.1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*. Vol 13(3), pp.341-360

Seiler, M. 2000. The efficacy of event study methodologies: Measuring EREIT abnormal performance under conditions of induces variance. *Journal of Financial and Strategic Decisions*, 13 (1), pp. 102-112.

Sharpe, W. 1991. The Arithmetic of Active Management. The Financial Analysts Journal. Vol. 47, No. 1.pp. 7-9

Sharpe, F. 1966. Mutual Fund Performance. *The Journal of Business*, 39, pp. 119-138.

Sharpe, W. F. 1992, Asset Allocation: Management Style and Performance Measurement, *Journal of Portfolio Management*,18, pp. 7-19.

Shi, W., & Seiler, M. 2002. Growth and value style comparison of U.S. stock mutual funds. *American Business Review*, 20, pp. 25-32.

Sullivan, R., Timmermann, A., and White, H.1999. Data Snooping, Technical Trading Rule Performance, and the Bootstrap. *Journal of Finance*, 54, pp.1647-1691.

Theia. 2020. Investment Management in the UK 2019-2020. Available from: https://www.theia.org/sites/default/files/2020-09/20200924-imsfullreport.pdf.

Theia. 2021. The Investment Association Annual Survey. Available from: https://www.theia.org/sites/default/files/2021-11/IA%20-%20Investment%20Management%20Survey%202020-2021.pdf. [Accessed on 03/03/2022].

Teo, M., and Woo, S. 2000. Persistence in Style-Adjusted Mutual Fund Returns. *Harvard University - Department of Economics.* Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=291372.

Treynor, L., Mazuy, K. 1966. Can mutual Funds Outguess the Market?. *Harvard Business Review*, 44, pp.131-136.

Treynor, L. 1965. How To Rate Management of Investment Funds. *Harvard Business Review*, 43, pp. 63-75.

Ter Horst, J., Nijman, T., De Roon, F. 2004. Evaluating style analysis. *Journal of Empirical Finance*, 11, pp. 29–53.

Theia: The Investment association. 2020. Investment Management in the UK 2019-2020. Available from: https://www.theia.org/sites/default/files/2020-09/20200924-imsfullreport.pdf.

Tonks, I. 2005. Performance persistence of pension-fund managers. *The Journal of Business,* 78 (5), 1917–1942.

Vysniauskas, P., and Rutkauskas, A. 2014. Performance Evaluation of Investment (Mutual) Funds. *Business: Theory and Practice,* 15(3), pp. 398-407.

Wermers, R. (2000). Mutual Fund Performance: An empirical decomposition into stock picking talent, style. *Journal of Finance*, 55 (4), pp. 1655-1695.

White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica. Vol. 48, pp. 817-838

Table A3.1. Data on UK unit trusts studies:

Authors/Year	Time span	Fund Covered	Returns	Database
Fletcher 1997	1981-1989	120 funds chosen at random from the universe of trusts.	Monthly returns/ offer prices	Money Management Magazine, and DataStream
Blake & Timmermann 1998	1972-1995	2375 funds (1402 surviving and 973 dead funds)	Monthly returns/ bid prices and net income (not include transaction costs or management fees.)	Micropal Ltd
Quigley & Sinquefield 2000	1978-1997	473 funds	Monthly returns/ bid prices and net income (TERs 1.35%)	S&P Micropal, investment objective, annual, charge are obtained from Unit Trust Yearbook
Fletcher &Forbes 2002-2004	1982-1996	724 funds (no survivorship requirements)	Monthly returns/ offer prices and net income.	Finstat managed fund database, Money Management, and Unit Trust Yearbook
O' Sullivan 2006	1975-2002	1,620 domestic equity funds (626 surviving and 216 dead)	Monthly returns	Standard & Poor's Database, investment objectives are obtained from (IMA)
Byrne, et al. 2006	1988-2002	421 funds tracked to the end of 2002, only 74 with continuous returns	Monthly returns/ offer prices	1988 Unit Trust Yearbook

Table A3.1. Continued.

Authors/Year	**Time span**	**Fund Covered**	**Returns**	**Database**
Gregory and Whittaker 2007	1989-2002	32 Ethical funds characteristic matched with 160 conventional funds	Monthly returns	S&P Micropal and Datastream
Cuthbertson, et al. 2010	1988 -2002	842 (actively managed funds	Monthly returns	Standard & Poor's Database, DataStream
Foran, et al. 2017	1997-2009	1,141 actively managed UK equity unit trusts and including 672 non-surviving funds	Monthly returns/ net of management fees	Morningstar