Benchmark Indices, Alpha creation and performance persistence

IRINA BEZHENTSEVA MATEUS

The thesis submitted is partial fulfilment of the requirements of the University of Greenwich to the degree of Doctor of Philosophy

August 2017

DECLARATION

I certify that the work contained in this thesis, or any part of it, has not been accepted in substance for any previous degree awarded to me, and is not concurrently being submitted for any degree other than that of "Doctor of Philosophy (PhD)" being studied at the University of Greenwich. I also declare that this work is the result of my own investigations, except where otherwise identified by references and that the contents are not the outcome of any form of research misconduct

Candidate: Irina Bezhentseva Mateus

Supervisors:
Dr. Alex Stojanovic

Dr. Natasa Todorovic

August 2017

ACKNOWLEDGEMENTS

This work is dedicated to my great husband and son to whom I was blessed to give birth during my PhD. I love you both very much and thank you for being in my life and for giving me your full support, the reason to progress and the strength to carry on during sometimes difficult times. A special word is to my mum for her unlimited love, understanding and constant unconditional support - both emotionally and financially, and thanks to both of my parents for raising me, believing in me and encouraging me to study abroad. You four are the most important people in my life and this thesis would not have been possible without you.

I would also like to take this opportunity to thank Dr. Natasha Todorovic, my supervisor, for her advice, guidance and collaborative work throughout this PhD period; my viva examiners Dr. Elena Kalotychou (Cass Business School, City University of London) and Dr. Gabriella Cagliesi (University of Greenwich) for their very helpful comments and suggestions and the chair Dr. Lesley Catchpowle.

Special and very sincere gratitude is reserved for Gillian Haxell for the best research student support. She was the one who helped me with all administrative problems which I faced during this PhD and was always available to provide useful and constructive advice in respect to the fees, all documents submissions and requirements.

ABSTRACT

This dissertation provides three self-contained empirical studies which investigate the role of benchmark indices, alpha creation and performance persistence. In the first essay, we re-visit the performance of 887 active UK equity mutual funds due to the fact that recent academic literature documents that standard benchmark models, such as FF3 and Carhart four factor models, produce economically and statistically significant non-zero alphas for passive benchmark indices. We use a new approach proposed by Angelidis, Giamouridis, and Tessaromatis (2013) which adjusts the alpha of a fund by the benchmark's alpha and, thereby, allows eliminating upward/downward biases in performance assessment caused by embedded benchmark alphas. In addition to the US evidence we identify persistently negative alphas of FTSE 100 Index in the period 1992–2013. By applying AGT method, we eliminate bias inflicted by benchmark alphas. The results show that adjusted Fama–French and Carhat alphas of UK equity mutual funds are higher than those implied by the standard three- and four-factor models and are overall positive.

The second essay re-visits the question of benchmark misclassification among 1281 US equity mutual funds and estimate its impact on benchmark-adjusted fund performance and ranking. All funds report S&P500 index as a prospectus benchmark, yet 2/3 of those are placed in the Morningstar category with risk and objectives different to those of the S&P500 index. We identify 'true' benchmarks for those mismatched funds and find that their S&P adjusted alphas are higher than 'true' benchmark adjusted alphas in 61.2% of the cases. In terms of fund quartile rankings, 30% of winner funds lose that status when the prospectus benchmark is substituted with a more suited one. In the remaining performance quartiles there is no clear advantage of using S&P 500 as a prospectus benchmark. The prospectus benchmark therefore can mislead investors about fund's relative performance. This leads us to conclude that any reference to performance in a fund's prospectus should be treated with caution.

In the third essay we assess UK mutual fund performance from a perspective of a peer-group, applying a novel approach suggested in Hunter et al. (2013). Our sample comprises of 817 UK long-only active equity mutual funds allocated to nine Morningstar style category peer-groups in the period 1992-2016. Overall, we find that those funds with most significant positive peer-group adjusted alphas continue to perform well one-year-ahead, using both parametric and non-parametric measures of persistence in performance. Further, a small increase in significance of

peer-group adjusted alphas significantly improves probability that a fund will be placed in the top quartile in the following period. Finally, we document that persistence in performance is driven by both winner and loser funds. The results within each peer group by and large conform to these findings.

CONTENTS

TABLES AND FIGURES

Chapter 3

Chapter 4

Chapter 5

Appendix

Chapter 1

Introduction

In the academic world the Capital Asset Pricing Model, the Fama and French three-factor model and the Carhart four-factor model have been adopted as standard benchmarks for performance evaluation. These models have been widely utilised to estimate the risk-adjusted performance of mutual funds. However, these three different approaches can lead to very different inferences and as it has been referred by different researches produce nonzero alphas even to passive benchmark indices. For example Cremers, Petajisto and Zitzewitz (2012) regressed the S&P 500 index on the Carhart four-factor model and obtained an annual alpha of 0.82 percent, statistical significant at 1 percent level over a sample period from 1980 to 2005. Considering the significant number of individuals who delegate portfolio management to active mutual funds it is important to understand and be able to evaluate the 'true' performance of fund managers according to the risk undertaken; therefore, it is essential to adjust fund performance for positive/negative alphas of passive benchmark indices.

From a practitioner/investor point of view the performance of active fund managers is commonly evaluated by comparing fund returns to returns of passive benchmark indices otherwise known as reference benchmarks (for instance, S&P 500 and FTSE 100 are used for large-cap stocks and the Russel 2000 and FTSE SmallCap for small-cap stocks in the US and the UK respectively). Reference benchmark can also be used to estimate relative fund performance in order to rank top and bottom performing funds with similar investment style. Several studies (for instance Angelidis et al.,2013) claim that self-reported prospectus benchmark is the most suitable reference benchmark for performance evaluation as it is better fits the investment fund strategy and objectives, therefore, it contributes to more precise inferences on fund manager performance. Research findings of the Investment Company Institute show that 34% of fund investors consult the fund prospectus before purchasing a mutual fund. However, recent evidence shows that self-reference benchmarks are often mismatched; moreover, studies claim that it can be done for strategic reasons. Thus, Cremers and Petajisto (2009) provide evidence that mutual funds typically have a high proportion of holdings that differ from those of the fund's theoretically correct benchmark index. Sensoy (2009) affirms that funds frequently differ significantly from their benchmarks. The results show that value funds are more likely to have self-designated benchmarks that are mismatched on value/growth, small-cap funds tend to have prospectus benchmarks mismatched on size. DiBartolomeo and Witkowski (1997) shows that return patterns of 40 percent of funds in the sample deviate from the benchmark declared in the prospectus with 9 percent of funds being seriously misclassified, two or more risk tiers away from their declared categories. Considering investors' close scrutiny of fund performance, it is important to investigate to which extent benchmark selection may affect the inferences on mutual fund performance; as conclusions drawn by investors based on the self-declared prospectus benchmark may be misleading. Moreover, it is crucial to investigate whether unsuitable benchmark selection may affect inferences on fund relative performance. Thus, for more accurate investor decision making fund ranking/performance in respect to peers should be re-examined in order to reveal whether winning funds are in fact winners and worst performing funds are truly losers.

Until recently the evidence of self-reported benchmark mismatches was documented only for the US. However, the last FCA report on the UK asset management market (FCA, June 2017) revealed that the issue of unsuitable benchmark selection may also be present in the UK. The report infers that category benchmarks may potentially misrepresent fund performance, and more should be done to clarify their use. On the other hand, even if equity mutual funds aim to disclose/allocate accurate benchmark alongside their historical returns, it is not always possible. For instance, there is no the off-the-shelf style specific index by a standard index provider (FTSE, MSCI) which can ,for instance, represent the style of unit trusts investing in small-cap growth stocks. The same is true for all other value and growth style categories. According to information provided by Morningstar database we account that 2/3 of funds facing this problem and currently report FTSE All Share Index as their benchmark.

Based on the above, this dissertation entitled *Benchmark Indices, Alpha creation and performance persistence* will be composed of three empirical essays, where in first (chapter 3), we will revisit the performance of UK mutual funds by adjusting fund alphas for non-zero passive benchmark alphas with Angelidis et al. (2013) approach. In second (chapter 4), we will estimate to which extend inaccurate benchmarking in the US mutual funds affects fund performance, its impact on fund relative performance and ranking. In the last essay (chapter 5), we will reexamine existing evidence on UK fund performance persistence with Hunter et al. (2014) methodology, which estimate the performance of a fund in excess the performance of the peer group; so that allows eliminating biases caused by abnormal benchmark indices and inaccurate benchmarking.

Chapter 3 estimates fund performance before and after the Angelidis, Giamouridis and Tessaromatis (2013) adjustment, examines the impact of style investment on value creation and controls for the fund performance in different market conditions (bull and bear market). The data set comprises 887 actively managed equity mutual funds with UK investment focus. The period of analysis spans from January 1992 to October 2013. Our results document non-zero alphas of a passive benchmark index FTSE 100. In contrast to the US evidence our findings indicate a negative annual benchmark alpha of -1.12 percent for the Fama and French three-factor model and the annual alpha of -1.13 percent for the Carhart four-factor model, statistically significant at one percent level. In addition, we show that the benchmark index alphas vary in accordance to different market conditions. Thus, in bear market benchmark alphas are significantly larger than in bull market, results range between -1.61 and -2.86 percent versus -0.47 and -1.10 percent, respectively. These results confirm that the standard Fama and French and Carhart models amplify the underperformance of UK equity mutual funds. The Angelidis et al. (2013) adjustment for the negative alphas in the benchmark index reveals, that UK focused equity funds are able to deliver positive excess performance. As example, for the whole sample period and the whole sample of funds AGT-adjusted Fama and French alpha exhibits a tenfold increase from just 13.81 bps (standard FF alpha) to 143.64 bps per year. The adjustment brings greater increase in alphas in bear rather than in bull market periods. For instance, the financial crisis period of 2008–2009 bares the adjusted Fama–French annual alpha which is 2.89% higher than standard alpha for the sample of our funds. This evidence is very important for academic literature as well as individual and institutional investors, as hitherto has been no evidence documenting positive outperformance of UK mutual funds.

Further in chapter 3 we estimate UK fund performance across different fund styles, where we split the funds into nine style categories according to Morningstar style-box. The results show, that after adjustment for negative benchmark alphas, UK fund performance across all styles shifts upward: negative alphas becomes less negative and in some cases turn to positive (albeit mostly insignificant), positive standard alphas notably increase and become significant. Among other findings we highlight that over 70% of mutual funds concentrate their portfolios in Small/Value, Small/Growth and Small/Blend stocks. This strategy pays off with the highest performance compared to other styles (generated positive adjusted FF3 alpha of 1.62%, 2.03%

and 1.53%, respectively; statistically significant at one percent level). In these style groups, positive abnormal performance persists even during market downturns. During the time period analysed Small/Value and Small/Growth funds provided the best performance and generated positive adjusted alphas in four out of five sub-sample periods. Partly, the results can be explained by the dot.com boom. Large/Value funds showed better performance than any other group during the financial crisis of 2008–2009. As robustness tests we replicated the analysis for a sub-sample of small capitalisation funds using FTSE Small Cap as a benchmark. Results confirm our previous findings and show even stronger significance for both, adjusted alphas and the differences between altered and standard alphas. Overall, our study concludes that non-zero benchmark alphas documented in standard Fama and French and Carhart models significantly bias the inferences on mutual fund performance and should be addressed. Thus, for the UK market, negative benchmark alphas significantly downgrade performance of mutual funds. It is the first study that accounts for these biases and provides evidence that UK focused equity funds are able to deliver positive excess performance, which is better than previous UK evidence suggests.

In chapter 4 we also apply Angelidis, Giamouridis and Tessaromatis (2013) methodology to eliminate the upward/downward biases in performance assessment caused by embedded benchmark alphas. Our sample consists of 1281 actively managed US equity mutual funds, for which we obtained net monthly returns from January 1992 to February 2016. According to the prospectuses all equity funds state S&P500 as Prospectus Benchmark. For comparative purpose we use Morningstar category classification (Global category), where funds are grouped into categories according to their actual investment style. Thus, for each mutual fund we identified a ‘true’ benchmark, the one more appropriate in accordance to Morningstar Global Category. After matching we realised that only 460 of the funds in our sample belong to the Large Cap blend Morningstar category, for which the S&P 500 would be the most suited benchmark. All other remaining funds fall across 21 other distinct Morningstar Global categories, some of which are with fund risk profile and composition very different from that of their prospectus benchmark. We estimate the impact of inaccurate benchmark allocation on fund performance and ranking using the largest five Morningstar categories S&P 500, Russel 1000 Value, Russel 1000 Growth, Russel Midcap and Russell 2000 index, which account for 80 percent of the sample (1,029 funds of a total 1,281). The remaining indices and their corresponding categories in our sample are not used for this analysis as the number of funds per category is not large enough resulting in some sub-periods featuring very few funds, jeopardising the objectivity of the results. Moreover, the remaining 20 percent of the funds are sector specific or country/region specific and call for sector or regional benchmarks. The performance assessment in this study is conducted with the Fama and French three and five factor and the Carhart four factors models as the standard models most widely used for performance evaluation.

Preliminary results confirmed that Global category benchmarks, which we call 'true' benchmarks provide a better fit for the estimation of mutual fund performance with on average 10 percent higher R-squared versus Prospectus benchmarks. Similarly to Cremers at al.,(2012) we document presence of non-zero alphas in passive indices when benchmark returns for S&P 500/'true' global category benchmark regressed on the excess returns of mutual funds. Thus, we report 33.01 bp annual alpha for S&P500, 74.93 bp for Russel 1000 Growth, -12, 58 bp for Russel 1000 Value, 60.17 bp for Russel Midcap and -197.01 for Russel 2000.

Analysis of fund performance estimated versus both prospectus and 'true' benchmarks for all mutual funds with 36 month rolling window showed that the results vary depending on the model and the reference benchmark applied. Overall, for the total sample period we document negative statistical significant adjusted alphas for FF3, FF5 and the Carhart models; where results estimated versus S&P 500 benchmark are better/overstated for Carhart and FF5 models (lower negatives alphas), and worse, higher negative adjusted alpha, when regressed with FF3 model . Fund performance estimated with the Carhart model for each mutual fund with 36 month period revealed that in 70 percent of the periods analysed the average alphas with S&P500 are higher and overestimate fund performance. Overall figures for the entire period show that 61.2 percent of the funds benefit from wrongly benchmarking their performance against S&P500 (prospectus benchmark). Thus, the average AGT-adjusted alpha drops by 23 basis points when a 'true' global categories benchmark is used. The results persist when calculated with and without overlapping periods.

Analysis of quartile ranking confirmed the importance of benchmark choice when fund performance is compared to peers. Thus, we show that 30 percent of the top performing funds move their quartile when performance is assessed with the 'true' benchmark. The difference in prospectus and 'true' benchmark adjusted alphas revealed that in 55 percent of the periods winners benefit from using S&P500 as a benchmark. Thus, in 2007-2009 and 2008-2010 such difference spreads to 460 and 320 basis points, respectively. Overall, the average advantage of stating S&P500 as the prospectus benchmark gives 68 basis points advantage to the winning funds. This leads us to conclude that strategic benchmark selection appears to be most likely in the funds at the top performance quartile, while we do not observe clear advantage of benchmark gaming in the remaining quartiles. Overall, we can highlight that the average alpha when the performance is adjusted with 'true' Global category benchmarks falls 28bps and 25bps in Quartiles 2 and 3, and increases 33bps in Quartile 4 in the whole sample period. The latter result shows that Quartile 4 funds get penalised from inaccurate benchmark selection. Thus, close to 30 percent of losers move up their quartile when performance is estimated with the most suitable global category benchmark. Hence, the choice of the benchmark affects not only the inferences about a fund's absolute performance, but it can also mislead investors about its relative performance. This leads us to conclude that any information in fund prospectus about the performance relative to the prospectus benchmark or any other funds should be treated with caution.

In chapter 5 we reassess the performance of UK mutual funds with active peer benchmark (APB) proposed by Hunter et al. (2014). This approach enables us to estimate fund performance/manager skills which are above the commonalities in fund strategies within a peer group and identify funds with best risk-adjusted performance within a group. Most importantly, it eliminates possible biases caused by abnormal benchmark alphas and inaccurate benchmark allocation. In this chapter we estimate fund performance and contract the ranking of the best and the worst performing funds. Based on the results we examine whether UK fund performance persists one year ahead and whether Hunter et al., (2014) methodology can serve as a good predictor of future performance.

To conduct this analysis we utilised a sample of 817 active UK long-only equity mutual funds in the period January 1992 to February 2016. Our results are in line with Hunter et al. (2014) and show that APB adjusted Carhart model has higher R-squared and explains UK mutual fund returns better than the standard one. The alphas of the two models are different. APB adjusted alpha is more statistically significant and is higher than Carhart alpha in 55% of the cases. Parametric (regression) and non-parametric (contingency tables) tests on performance persistence revealed that APB adjusted alpha is a strong predictor of performance one-year-ahead when tested in the UK. It shows that 1% increase in t-statistics of APB-adjusted alphas leads to 2.37% increase in probability that a fund will be placed in the top performing quartile. Moreover, 1% increase in APB adjusted alpha will have 15.53bps higher four-factor alpha one year ahead. Our results document fund performance persistence for the UK. Moreover, the performance persists regardless of the method employed. Thus, we conclude that the investors selecting the funds with highest peer-group-adjusted alphas will generate statistically and economically significant higher excess returns/alphas one-year ahead. Results remain robust either we divide funds into four performance quartiles or split them into deciles. Our findings reveal that persistence is driven by both winner and loser funds. The result is consistent across Morningstar peer-groups. Previous evidence in the UK only documented persistence for poor performers.

This dissertation is composed of chapter 1 introduction, chapter 2 literature review covering the existing literature in the area of research, chapter 3,4 and 5 presents empirical essays, chapter 6 concludes. Each empirical essay in this dissertation will include a section with an introduction and existing literature review, a section where we describe the data and methodology, a section which explains the main and, in some chapters, preliminary results and the final section, which concludes the chapter.

Chapter 2

2 Literature review

2.1 Asset pricing models

The first evolvement of the modern portfolio theory began with Markowitz (1952) mean-variance model. The model was designed to build the optimal portfolio and explained how for a given level of risk is possible to achieve maximum level of expected return, or equivalently minimize level of risk for a given expected return using diversification. The work of Sharpe (1964) and Lintner (1965) continued the model and constructed the Capital Asset Pricing Model (CAPM). The model revealed that differences in expected returns across securities and portfolios are due to the systematic risk which cannot be eliminated through diversification. Therefore securities and portfolios more exposed to systematic risk are expected to provide higher returns in excess to the risk free rate. As a result, investors can construct portfolios relative to their risk preferences by choosing the stocks according to its value of market beta.

The CAPM was widely accepted and became very famous even though had known shortcomings. One of the most recognised critics belongs to Fama and French (1993) who stated that the empirical record of the model is poor; it has many simplifying and unrealistic assumptions such as an unrestricted risk-free borrowing and lending as well as the investor's homogeneous expectations about expected returns, among others.

Based on the previous literature of Banz (1982) who noticed that market size (ME) adds to the explanation of the cross-section of average returns provided by market beta; Bhandari (1988) who described the positive relation between leverage and average return and explained that leverage helps define the cross-section of average returns in tests when market beta and size (ME) are included; Chan, Hamao and Lakonishok (1991) who noticed the role of book-to-market equity (BE/ME) in explaining the cross-section of average returns on Japanese stocks; Ball (1978) who states that E/P ratio is a "catch-all proxy" for different factors in expected returns that could be applied to size (ME), book-to-market equity and leverage which state that size and book-to-market equity Fama and French (1993) hypothesized that the best approach to determination of the risk in the market is a multi-index model where size, leverage, earnings/price ratio and book-to-market equity are average return variables.

The study of the joint roles of those variables in the cross-section of average stock returns revealed that beta alone or in combination with other variables provides little information about average returns. The combination of such variable as size and book-to -market equity absorbed the roles of E/P and leverage in average returns. Fama and French (1993) found that book-to-market equity and size are related to economic fundamentals. For instance, firms with high BE/ME are inclined to have low earnings on assets whereas the opposite tendency, high earnings, was observed for low BE/ME firms. In addition, it was highlighted that size is associated with profitability and small companies are likely to have lower earnings on assets compared to large ones. Authors claim that book-to-market equity and size are proxies for distress. Thus, small firms can be exposed to long earnings depression that circumvents big firms. They stated that profitability is a source of common risk factor in returns and it is also indicated by the positive relationship between BE/ME and average return.

Based on the above, Fama and French (1993) posit that size and book-to-market equity play a significant role in the performance of the stock and do a good job explaining differences in average stock returns. Authors proposed the three-factor model comprised of a market factor, size (SMB) and book-to-market equity factor (HML) as proxies for risk which captures patterns in returns known to cause problems for the CAPM. Thus, SMB (small minus big) factor is a proxy for risk in returns associated with size. It is the difference in returns of small- and large-stock portfolios; HML (high minus low) factor is a proxy for risk which captures value premiums in returns. It is built as the difference between the returns of a portfolio of high-book-to-market and low-book-to-market (BE/ME) securities. As it is explained in Griffin and Lemmon (2002), it captures premiums in returns focus on risk and investor overreaction. Market factor (RM-RF) is the excess market return, in other words excess return on a broad market portfolio.

Hence, the Fama and French three factor model explains that the expected return on a portfolio in excess of risk-free rate depends on the sensitivity of its return to the three factors; to illustrate, excess return of portfolio *i* can be estimated using the following equation:

$$R_{i,t} - R_{F,t} = \beta_{i,M}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{i,SMB}(R_{SMB,t}) + \beta_{HML}(R_{HML,t}) + e_{i,t} \quad [1]$$

where $R_{i,t}$ is the return of a portfolio i in period t, $R_{F,t}$ is the US 1 month Treasury bill (risk free rate), $R_{M,t} - R_{F,t}$ is the market risk premium (US market risk premium is defined as the value-weighted return of all CRSP stocks incorporated in the US and listed on the NYSE, AMEX, or NASDAQ (R_M) minus one month US Treasury bill rate (R_F)[1]), *SMB* and *HML* are size (small minus big returns) and value (high minus low book-to-market returns) factor portfolios (R_{SMB} and R_{HML}), respectively. $\beta_{i,M}$, $\beta_{i,SMB}$, β_{HML} are factor loadings, and $e_{i,t}$.is the error term.

To form SMB and HML factor portfolios (R_{SMB} and R_{HML}) stocks were sorted by market capitalization into two size groups, Big (B) and Small (S) portfolios, with the 50 percent size breakpoint and then sorted by book-to markets into three value groups, High (H), Medium (M) and Low (L), with 30/40/30 breakpoint. The sorts were performed independently of each other, and their intersections were used to build six value-weighted portfolios, S/L (return on stocks that are in the Small portfolio and the Low book-to market portfolio: $R_{S/L}$), S/M ($R_{S/M}$), S/H ($R_{S/H}$), B/L ($R_{B/L}$), B/M ($R_{B/M}$) and B/H ($R_{B/H}$).

Thus, SMB factor portfolio (R_{SMB}) was constructed as the equal-weight average of the returns on the small stock portfolios minus the returns on the big stock portfolios: $R_{SMB} = ((R_{S/L} - R_{B/L}) + (R_{S/M} - R_{B/M}) + (R_{S/H} - R_{B/H}))/3$. Similarly, HML factor portfolio returns (R_{HML}) were defined as the equal-weight average of the returns on the value stock portfolios minus the returns on the growth stock portfolios: $R_{HML} = ((R_{S/H} - R_{S/L}) + (R_{B/H} - R_{B/L}))/2$.

Fama and French (1993) argue that the three-factor model capture many of the CAPM average-return anomalies and can be used as a good measure of returns of portfolios formed on size and BE/ME. Fama and French (1996) explained that the strong patterns in returns were also observed regarding the portfolios constructed on E/P, C/P (cash flow/price), and sales growth. As it was explained, stocks with high earnings/price, high cash flows/price and low sales growth usually have high BE/ME that tend to load positively on HML. These stocks are relatively distressed and generate higher average returns. Inverse relationship can be described for stocks with negative slopes on HML.

[1] from Ibbotson Associates

The Fama and French three-factor model became widely accepted and implemented; however CAPM model is still very popular among investors and analytics due to its simplicity. The three-factor model was examined and also critically discussed by different researches. Many of them confirmed that the model can explain considerable variation in returns but there are others who argue that there are areas for improvement.

The major improvement to the three-factor model was made by Carhart (1997) who addressed the issue of momentum anomaly. His research was based on two papers Fama and French (1993) and Jegadeesh and Titman (1993). Jegadeesh and Titman (1993) explained the momentum effect and stated that stocks that performed well in the past would continue to perform well in the future. Particular if investors hold stocks that delivered high returns in the past and sell the ones with poor performance will be able to generate significant positive returns over 3- to 12-month holding period. Carhart (1997) criticized the statement and explained that this impact should be driven by fact that mutual funds tend to have relatively larger positions in last year's leading stocks and cannot be explained by momentum strategies or superior stock-picking skills. In addition, he disagreed with Grinblatt, Titman, and Wermers (1997) who claimed that mutual funds using momentum strategies can generate higher returns before management fees and expenses. To check these results he constructed four-factor model based on three-factor model from Fama and French and added an additional factor capturing one-year momentum effect. Momentum factor (WML) was built as the difference between the returns on diversified portfolios of winners and losers.

$$R_{i,t} - R_{F,t} = \beta_{i,M}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{i,SMB}(R_{SMB,t}) + \beta_{HML}(R_{HML,t}) + \beta_{WML}(R_{WML,t}) + e_{i,t} \qquad [2]$$

where WML factor portfolios (R_{WML}) were constructed as the difference between the equal-weight average of firms with the highest 30 percent 11-month returns lagged one month (winners) and the equal-weight average of firms with the lowest 30 percent 11-month returns lagged one month (losers). The rest of the equation is equal to the Fama and French (1993) model above. The portfolios in the analysis were comprised of all NYSE, AMEX, or NASDAQ stocks and were re-formed monthly.

The tests of the model provided in Carhart (1997) showed that the Fama and French three factor model is superior to CAPM since size and book-to-market equity factor are included. However, four-factor model substantially improves on the average pricing errors of the CAPM and the three-factor model.

2.2 The Alpha

The Fama and French three-factor model and the Carhart four-factor model became very famous and got wide acceptance and implementation among investors, fund managers and analytics. They serve as standard benchmarks for performance evaluation and performance persistence and help in relative assessment, when performance of funds relative to peers is evaluated.

Arguably, there's one question that all investors want the answer to. "How can I make higher returns?". With this target, significant number of individuals delegate portfolio management to active mutual funds due to their prior belief in managerial skills, experience and the ability to beat the market. Thus, according to Morningstar, since 2007 assets under management held in active funds have grown 54 percent globally, to 24 trillion dollars (Monney, 2016).

However, there is another option which allows investing in passive mutual funds which offer lower-cost exposure to markets by tracking an index. To make a decision on where to invest and which type of funds offer better value for money investors employ standard benchmark models such as Fama and French three and the Carhart four factor models which allows evaluating the ability of active mutual funds to generate risk-adjusted excess returns.

The risk adjusted excess return represent portfolio manager's forecasting ability/stock picking skill, which allows to generate realised returns on the market portfolio above the ones predicted by the standard models. One of the first and most famous papers which aimed to evaluate portfolio manager's "predictive ability" to earn excess return for the given level of riskiness of the portfolio was written by Jensen (1968). The study explained that if a portfolio manager has forecasting ability and managerial skills then she should be able to generate abnormal return in excess of the one predicted by the market. Therefore, if it is the case it should be seen in the positive intercept (alpha) after the performance of his/her portfolio is estimated with the CAPM model for the given portfolio's beta and the average market return. Thus, zero or negative alpha means that a portfolio manager does not possess such managerial skills or forecasting ability, is unable to beat the market and simply replicate the market, while negative performance indicates unsuccessful forecasting ability and generation of too high expenses.

With the development of the Fama and French three-factor and the Carhart four-factor models it is common to estimate alpha (performance of a fund/portfolio) as the return in excess of exposure to three and four risk factors (market, size and value factors, including momentum factor for the Carhart).

Thus, the aforementioned three and four-factor models can be rewritten as follows:

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{i,M}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{i,SMB}(R_{SMB,t}) + \beta_{HML}(R_{HML,t}) + e_{i,t} \quad [3]$$

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{i,M}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{i,SMB}(R_{SMB,t}) + \beta_{HML}(R_{HML,t}) + \beta_{WML}(R_{WML,t}) + e_{i,t} \quad [4]$$

where α_i represents the excess return of the portfolio *i* over period *t*. Intercept α_i should be equal to zero for all securities and portfolios if the factor exposures (betas) β_M, β_{SMB}, β_{HML}, β_{WML}, capture all variation in expected returns. The rest of the equations is as per the Fama and French (1993) and the Carhart (1997) models explained above.

2.3 Empirical evidence

2.3.1. International evidence

During the last three decades many researches have provided different empirical evidence using the standard benchmark models. Many academics constructed test portfolios based on the three-factor model specification in order to verify its validity and accuracy. The evidence has been shown at the domestic and international levels. Significant number of academic papers documented results validating the Fama and French's model.

For example Arshanapalli, Coggin and Doukas (1998) estimated Fama and French three-factor model by using international portfolio returns from 18 equity markets and stated that the model can justify why average industry stock returns are above the risk-free rate. The results showed that the three-factor model capture the variation in average industry returns. "The lowest R^2 is 0.66, while in the most regressions it is greater than 0.71", so that shows that even portfolios are split by industry the results are consistent with the empirical evidence of Fama and French (1996) and can be utilised to provide a good explanation of international industry returns. Liew and Vassalou (2000) constructed portfolios using the data from 10 countries and found that HML and SML contain substantial information about future GDP growth. According to the results, the returns of HML are higher when portfolios are periodically rebalanced. They also stated that momentum is very sensitive to the rebalancing period. Longer rebalancing intervals negatively affect returns to momentum effect. In addition, they confirmed the existence of the size premium. The results showed that in all countries, except Switzerland, SMB generates positive returns. The evidence is statistically significant in Canada, France, Japan, and the United States. Maroney and Protopapadakis (2002) examined seven countries from 1982 to 1994, confirmed positive relation between returns and book-to-market ratio and stated negative relation between returns and market value among all countries analysed. Hence they argue that the three-factor model has the international effect.

Faff (2001) used data from Australian stock market over the period 1991 to 1999 and found strong support that the three-factor model performs well with monthly data. Drew and Veeraraghavan (2002) found evidence on the size and value premium by employing the data from Malaysia for the period December 1992 - 1999. Using a multifactor model developed for the Malaysian setting they found that the two mimic portfolios, SMB and HML generate average annual returns of 17.7% and 17.69% per annum, respectively. They suggested that the size and value premium are the compensation for the risk undertaken which is not captured by the CAPM.

The Carhart (1997) model which added the fours momentum factor to the standard three-factor model also became widely accepted and widely used in the literature for modelling stock returns. Hence, the empirical work of L'Her, Masmoudi and Suret (2004) provided evidence in support of the four factor pricing model from the Canadian stock market. For the July 1960-April 2001 sample period they obtained the average annual premium for the market, size, book-to-market and momentum risk factors equal to 4.52, 5.08, 5.09 and 16.07% respectively. Similar results showed Liew and Vassalow (2000) for the 1976-1996 period. Lam, Li and So (2010) examined the application of the four-factor model in the Hong Kong and provided evidence that all four factors in the model are significant.Authors stated that reasonably high values of adjusted R^2 (0.68) and the insignificance of an additional explanatory variable of residual standard deviation provide good support to the model. Lozano (2006) compared Fama and French three-factor model and the Carhart four-factor model using a sample of 871 monthly observations and 25 test portfolios. He stated that the spread and magnitude of the alphas in the Carhart model are smaller. Also it produces a fewer number of statistically significant results. Different tests demonstrate that the Carhart model shows smaller pricing error, therefore, more accurate compared to Fama and French. Several tests rejected Fama and French but run properly with the Carhart. As the result, Lozano (2006) concluded that momentum definitely helps to improve the model to price the average returns on portfolios.

The Fama and French three-factor model was accepted by many researchers and practitioners internationally. However, there is a significant strand of academic literature that criticized the model and highlighted that it requires some improvements.

Significant strand of literature criticised the conjecture of Fama and French (1993, 1996) that the value and size effects are compensation for the risk of financial distress. Vassalou and Xing (2004) analysed the effect of default risk on equity returns and provided evidence that the size and the book to-market (BM) effects are default effects, which exist only within top quintiles with the highest default risk. Consequently, there are no size or BM effects in the remaining stocks of the market. In addition, the results show that when compared to low default risk firms high-default-risk firms earn higher returns only when they are small in size and high book to-market. In other circumstances, high-default-risk firms do not earn higher returns, even if their risk of default is actually very high. Based on the results, the study affirms that SMB and HML factors include some default-related information; however, it does not fully explain the cross section of equity returns.

Griffin and Lemmon (2002) provide similar evidence and state that book to-market effect is concentrated in high default risk stocks and is driven by the poor returns of low book-to market stocks rather than the superior returns of high BM stocks. They document large, more than twice, return difference between high and low book-to market stocks, which cannot be explained by the three-factor model. Daniel and Titman (1997) assert that most of the comovements of high book-to-market and small capitalisation stocks are not due to distressed stocks being exposed to a unique "distress" factor. They state that there is no evidence of a separate distress factor. In fact, according to the authors, stocks with similar factor sensitivities tend to become distressed at the same time. The results show that it is characteristics determine expected returns but not factor loadings. Hence, small and high book-to-market stocks just act as proxies for these characteristics. Therefore, the study claims that it is the characteristics rather than the covariance structure of returns that explain the cross-sectional variation in stock returns.

In opposite, the study of Bourguignon and Jong (2003) provided evidence in support of the hypothesis that the value premium is related to a distress factor rather than to stock characteristics. The paper of Gharghori, Chan and Faff (2007) emphasised that SMB and HML factors are not proxies for default risk, however they capture some forms of priced risk. The analyses conducted on the model's ability to explain equity returns revealed that the three-factor model does a good job in explaining equity returns and is vastly superior to the CAPM. An augmented version of the three-factor model with additional default-risk factor showed similar to the 'standard' model performance. As a result, authors highlighted that the improvement from the augmented model is marginal at best. Campbell, Hilscher and Szilagyi (2008) highlighted that since 1981 financially distress stocks have generated anomalously low returns. Nevertheless, their market betas, standard deviations, and loadings on the SMB and HML factors are much higher than stocks with a low risk of failure, which is in opposite to the Fama and French' statement that the value and size effects are compensation for the risk of financial distress. The results are consistent among all size quantiles, with a particular strong impact observed in smaller stocks.

Several academics questioned international implication of Fama and French three-factor model. Asness, Moskowitz and Pedersen (2013) construct a simple three-factor model that captures the global returns across asset classes. For this purpose they build value and momentum portfolios and examine it across four equity markets: the US, the UK, continental Europe and Japan. Evidence highlight significant return premium to value and momentum across all the markets. Griffin (2002) examine whether the global version three-factor model explains time-series variation in international stock returns better than country-specific (domestic) model. Results reveal that domestic models have higher explanatory power and, for the majority of tests performed, generate lower average pricing errors comparing to the world model-particularly for individual securities. Moreover, country-specific model yields lower out-of-sample pricing errors. It shows that performance evaluation estimated on a country-specific basis gives more accurate results.

Kosowski (2011) argue that unconditional risk-performance measures are biased in time of boom and recessions and underestimate the value added by actively traded mutual funds. Recent paper of Asness, Moskowitz and Pedersen (2013) examined the relationship between asset's return and value and momentum effects across eight diverse markets and various asset classes and documented consistent value and momentum return premia across all asset classes. They looked at the liquidity risk as a possible explanation. However, the results provides interesting facts: liquidity risk is negatively related to value and positively related to momentum (consistent with Pastor and Stambaugh (2003), Sadka (2006)) so this means it cannot explain why both value and momentum deliver positive return premium. They leave the question for the further research: why negative value loading on liquidity risk gives positive risk premia.

Bartholdy and Peare (2005) compared the performance of CAPM and Fama and French three factor model for individual stocks using CRSP data from 1970 to 1996. The results show that excess return is best explained when five years of monthly data is used together with the equal-weighted CRSP Index, in contrast to the commonly recommended value-weighted index. The study brings into question the use of either model for estimation of individual expected stock returns and shows that CAPM model can explain on average only 3% of differences in returns, while the Fama and French three-factor model is slightly better, with R^2 of around 5%.

A significant strand of literature identified anomalies which are not explained but known to cause problems to the Fama and French (1993) model. One of them is net share issues. Thus, Ikenberry et al. (1995) document large abnormal returns after stock repurchases announced by value stocks. The work highlights that firms buying back their stock overperform for a period of four years. Conversely, (Loughran and Ritter (1995) provide evidence of low returns after stock issues, in addition to Spiess and Affleck- Graves (1995) they show that firms making seasoned stock issues substantially underperform similar size matching firms that did not issue equity for a period of five years. Momentum and the post-earnings-announcement drift anomaly were discussed in Sadka (2006). Thus, Kim and Kim (2003) propose to augment Fama and French three-factor model by including an additional risk factor unexpected earnings surprise (ES), which, according to the results, substantially improves the standard three factor model in explaining post-earnings-announcement drift. The factor is constructed as the difference in returns on the positive earnings surprise portfolios and returns on the negative earnings surprise portfolios. The earnings surprise is measured as difference between actual earnings and the average of analysts' earnings forecasts that proxy for the market expectations.

Kothari and Warner (1997) provide results that tests for long-horizon (i.e., multi-year) abnormal security returns around firm-specific events are misspecified. Tests of a sample of 200 securities with the Fama and French three-factor model revealed abnormal performance over a 36-month horizon for 34.8% of the samples (two-tailed parametric tests at the 5% significance level). The study observes both positive and negative abnormal performance which persists following simulated events. It concludes, that considering that long-horizon tests are misspecified and that parametric test statistics do not satisfy the assumed zero mean and unit normality assumptions, the results of long-horizon studies should be treated with extreme caution. Nonparametric or bootstrap procedures were proposed to reduce such misspecification.

Ang et al. (2006) highlighted that volatility is another anomaly which is not explained by three-factor model, as the model does not account for either the low average returns earned by stocks with high exposure to systematic volatility risk or for the low average returns of stocks with high idiosyncratic volatility.

Accruals were called one of the most pervasive return anomalies left unexplained by the three-factor model. First time the issue was mentioned in Sloan (1996) who stated that firms with high accruals underperform firms with low accruals. Fairfield, Whisenant and Yohn (2003) described that accruals could be viewed as a general market mispricing of growth in long-term net operating assets. Later paper of Fama and French (2008) admits that net stock issues, accruals, and momentum cause serious problem to FF3 model with the asset growth and profitability anomalies being less robust. The anomalous returns associated with net stock issues, accruals, and momentum were documented in all size groups (micro, small, and big) in cross-section regressions and were also considered strong in sorts, at least in the extremes. The results showed that asset growth anomaly show up in average returns on microcaps and small stocks, and profitability anomaly (abnormal high returns) tends to be present among profitable firms with higher profitability.

More recently, the study of Novy-Marx (2013) claimed that that Fama and French's HML factor would be more profitable if it is constructed controlling for profitability. Author stated that gross profitability has much more power than earnings and is able to explain most earnings related anomalies, as well as a large number of seemingly unrelated anomalies. The study explains that strategies based on profitability are growth strategies, which therefore dramatically improve a value investor's investment opportunity set. Thus, gross profits-to-assets is complimentary to book-to-market and has more information above that contained in valuations. The study proposes to use augmented value and profitability factors such as HML/GP ("HML conditioned on gross profitability") and PMU/BM ("profitable-minusunprofitable conditioned on book-to-market”). The results show that both new factors have a larger information ratio relative to the Fama-French factors plus UMD momentum factor.

Furthermore, Novy-Marx (2013) claims that industry-adjusted gross profitability has even more power than gross profitability predicting the cross-section of expected returns. The work proposes a four-factor model which consists of the market and industry-adjusted value, momentum and gross profitability “factors,” and affirms that the model performs remarkably well pricing a wide range of anomalies, including (but not limited to) strategies based on return-on-equity, market power, default risk, net stock issuance and organizational capital.

Aharoni, Grundy and Zeng (2013) analysed the relationship between expected profitability, book-to market, expected investment and expected returns. They criticised Fama and French (2008) and argued that the problem of their model is that it considers only the relation between current investment and expected returns, when in fact it should look at the relation between expected future investment and expected returns. Moreover, the study emphasized that for the correct estimation of the relation between expected returns and the three variables valuation should be conducted at firm level rather than per-share level. The results showed a positive relation between expected return and book-to market, a positive relation between expected return and expected profitability; however, underlined a negative relation between returns and expected investment. Similarly Chen, Novy-Marx and Zhang (2011) highlighted the role of investment in explaining the cross-section of returns and proposed a new three-factor model in which investment and profitability (return on assets) are the main explanatory variables. Based on the results the new model outperforms standard asset pricing models and is able to explain anomalies associated with net stock issues, asset growth, total accruals, earnings surprises and financial distress. Authors claim that newly proposed factors are built on economic fundamentals and tie expected returns to firm characteristics.

2.3.2. Changes to the models proposed

Such massive criticism led to the emergence of a vast literature aiming to propose augmented Fama and French and Carhart models for more accurate estimation of asset pricing performance.

Ferguson and Shockley (2003) argue that firm-specific variables that correlate with leverage (such as book-to market and size) will appear to explain returns after controlling for proxy beta due to the fact that they incorporate the missing beta risk. They propose to build augmented three factor model formed with relative leverage variable (debt-to-equity ratio) and relative distress (Altman's Z) and state that it will outperform the Fama and French (1993) standard model in the cross section. The results from Fama-MacBeth (1973) cross-sectional regressions showed that the augmented three-beta model has more explanatory power than the Fama and French (1993) three factor model, however, it did not worked as expected in time-series tests. As a result, authors confirmed that SMB and HML are very important in time series estimations and stated that maybe relative distress is indeed a priced factor. However, as another view they suggested that SMB and HML factors may enclose some information. For instance, according to Brennan, Wang, and Xia (2004) it may encompass information about the changing investment opportunity set, so the loadings on SMB and HML could measure sensitivities to the state variables. Alternatively, consistent with Berk, Green, and Naik (1999) a firm's book-to-market ratio may bear information about its changing risk (relative to its asset base), while size may convey the importance regarding growth options in relation to its assets in place.

The importance of incorporating liquidity as an additional risk factor to Fama and French three-factor model was also discussed by many researchers. For instance, Eckbo and Norli (2002, 2005), proposed to construct a liquidity factor as"low-minus-high" stock turnover portfolio; whereas Pastor and Stambaugh (2003) proposed to use order-flow related return reversals. Liu (2006) propose a liquidity measure as the standardized turnover-adjusted number of zero daily trading volumes over the prior 12 months which, according to the author, captures various dimensions of liquidity such as trading cost, trading quantity and speed. The work tests two-factor (market and liquidity) model. Although it states that Fama and French three-factor model have better explanatory power than the CAPM, it claims that the proposed two-factor model can explain better the performance of portfolios classifies by cash flow-to price, earnings-to-price and dividend yield.

Recent research of Foran and O'Sullivan (2014) suggested adjusting the Fama and French (1996) and the Carhart (1997) models by incorporating such liquidity factors as a stock characteristic (illiquidity level) and systematic liquidity risk for the best-fit model. To test it they build a liquidity factor mimicking portfolio, firstly, by using an illiquidity characteristic risk mimicking portfolio and then apply a systemic liquidity risk mimicking portfolio. Findings confirm the important role of liquidity as a stock characteristic and systematic liquidity risk in UK mutual fund performance evaluation. Moreover it shows a statistically significant shift leftward (reduction) in the cross-sectional distribution of the three factor alphas when control for stock holdings' liquidity. The results are also robust to a momentum factor for the majority of the sample. Otten and Reijnders (2012) examine the performance of mutual funds that invest in UK smaller companies and build a small cap version of the four-factor model where besides adding a liquidity risk factor (LMH) they include a dummy variable which controls for a January effect discussed previously by Keim (1983). Gharghori, Chan and Faff (2007) proposed to alter Fama-French model by including three additional variables: leverage, momentum and liquidity. However, the test on the ability of the augmented model to explain equity returns showed only a marginal improvement in the model's explanatory power.

Hou, Karolyi and Kho (2011) follow Fama and French (1992,1993) methodology and construct factor mimicking portfolios based on monthly returns of stocks from 49countries to estimate which factors are important for explaining the time-series and cross-section variation in global stock returns. For the analysis they use variables like cash flow/price, dividend/price, book-to-market equity, size, earnings/price, momentum and leverage that according to the international empirical asset pricing literature to be cross-sectionally correlated with average returns. Evidence show that the cash-flow-to-price-factor is more relevant price ratio for HML factors construction since it produces fewer model rejections than factors sorted on book-to-market ratios.

Moreno and Rodriguez (2009) claim that the effect of systematic skewness is important to consider for the analysis of mutual fund performance evaluation since it helps to explain the time variation of risk premiums. They add the coskewness factor as an additional variable to the CAPM and the four-factor model and state that it increases the explanatory power of the models due to the fact that the sign of the variation in the performance is determined by the loading on the coskewness factor.

Banerjee, Doran and Peterson (2007) examine the relationship between the future return and both current level and innovations of implied volatility. The results indicate significant relationship between future return and both VIX levels and innovations with even stronger results for high beta portfolios. As a result authors assert that VIX show predictive power for future returns and can be used as a priced risk factor.

Besides additional factors to the Fama and French three-factor model academic literature also proposed new models/methodologies which outperforms the standard models and suggest more accurate securities/portfolio performance evaluation. Kosowski et al. (2006) show that higher moments in individual find alphas as well as heterogeneous risk-taking among funds may lead to thick- or thin-tailed cross-section alpha distribution that points on a complex non-normal distribution in the cross section of mutual funds alphas. They argue that for the proper assessment of the managerial skills a bootstrap methodology has to be applied when analysing the cross section of t-statistics, which controls for the expected idiosyncratic variation in mutual fund returns. To test it they examine the performance and performance persistence of the "best" and "worst" funds based on a sample of the U.S. open-end, domestic equity mutual funds over the 1975 to 2002 period by using Carhart four-factor model with and without the bootstrap methodology applied. Authors highlight the importance of using bootstrap for the more accurate inferences, particular for smaller samples of funds (or shorter time series), among groups of funds with lower right-tail levels of performance, or at least rankings that use the appraisal ratio or the *t*-statistic of the alpha. Overall, the results show the evidence that sizable minority of managers among growth-oriented funds is able to deliver superior performance and performance persistence.

Blake et al. (2014) used new methodologies such as panel method with both fund and time effects as well as panel bootstrap methods for a more accurate estimation of mutual fund performance against benchmarks. In addition they add another factor such as fund size to the standard evaluation models. Moreover, they alter the four-factor model by adding fund size as an additional variable to control for fund-specific characteristic. The results show that average UK equity fund manager is unable to beat the benchmark after fees and expenses, even when they control for the best performing funds.

Gregory, Tharyan and Christidis (2013) test alternative forms of the Fama and French and Carhart models for the UK market. They construct portfolios based on the sample of the largest 350 firms by market capitalisation and form the portfolios and factors using the methodology described on Ken French's website. The work examines different methodologies using different approaches to factor construction. Besides the standard three and four-factor models it tests augmented versions with value-weighted factor components proposed by Cremers et al. (2012) and decomposed factors suggested by Zhang (2008), Fama and French (2012) and Cremers et al. (2012). Moreover, they check the ability of each model to explain the cross-section of returns in portfolios sorted on the basis of prior volatility The results show that, when factors are formed excluding smaller firms, value weighting and decomposing factors lead to a modest improvement in performance, where the value-weighted decomposed four-factor model seems to give a better estimate over the other models applied. Tests on the large firms sample demonstrate that all models provide a plausible explanation of the cross-section of UK returns even when portfolios are formed on the basis of momentum. However, authors raise doubts that risk factors are consistently and reliably priced since they vary with the test portfolios employed.

Hou et al. (2015) propose a new q-factor model consisting of the market factor, a size factor, an investment factor, and a profitability factor. The model tested Fama and French three-factor model and the Carhart four-factor model against 80 anomalies, which are grouped into 6 categories: (i) momentum; (ii) value-versus-growth; (iii) investment; (iv) profitability; (v) intangibles; and (vi) trading frictions. The study infers that many claims in the anomalies literature are exaggerated, likely by excessively weighting on microcaps. The new model proposed performs as well as, and often outperforms the FF3 and the Carhart models across major categories of anomalies (except for the operating accrual and R&D-to-market anomalies).

2.3.3 Fama and French five-factor model and empirical evidence

Significant criticism of Fama and French three-factor model and new evidence on the impact of profitability and investment on expected returns (such as Novy-Marx, 2013 and Aharoni, Grundy and Zeng, 2013) leaded to a new Fama and French five-factor model (Fama and French, 2015). The new model included five risk factors, the market, size and book-to market factors (as for the three-factor model) and additional profitability and investment factors, which, according to Fama and French, were chosen as natural choices as both, profitability and investment add to the description of average returns provided by book-to market.

The model was written as:

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_M\left(R_{M,t} - R_{F,t}\right) + \beta_{SMB}(R_{SMB,t}) + \beta_{HML}(R_{HML,t}) + \beta_{RMW}(R_{RMW,t}) + \beta_{CMA}(R_{CMA,t}) + e_{i,t} \qquad [5]$$

where $R_{i,t}$ is the return on security or portfolio i in period t, R_{Ft}, is risk free rate (US 1 month Treasury bill), R_{Mt} is the return on the value-weigh market portfolio (the value-weighted return of all CRSP stocks incorporated in the US and listed on the NYSE, AMEX, or NASDAQ), SMB and HML are Fama and French (1993) size (small minus big returns) and value (high minus low book-to-market returns) factors respectively, RMW and CMA are new profitability and investment factors calculated as the difference between the returns of stocks with robust and weak profitability (RMW) and the difference between the returns of low and high investment firms (conservative minus aggressive (CMA)), and $e_{i,t}$ is error term. Intercept α_i should be equal to zero for all securities and portfolios if the factor exposures (betas) β_M, β_{SMB}, β_{HML}, β_{RMW}, β_{CMA} capture all variation in expected returns.

The size and value factors are constructed using independent sorts of stocks into two Size groups and three B/M groups (HML) (2x3 factors). NYSE median market cap was applied as the Size breakpoint. The 30th and 70th percentiles of B/M for NYSE stocks were used for the B/M breakpoints. The profitability and investment factors of the 2x3 sorts, are built in the same way as HML (two Size groups and three OP groups (RMW), or three Investment groups (CMA). Operating profitability, OP, in the sort for June of year t was calculated using accounting data for the fiscal year ending in year t-1 and was estimated as revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by book equity. Investment was measured as the rate of growth of total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1.

The results revealed that the value factor becomes redundant for describing average returns when profitability and investment factors are added, as high average returns are fully captured by its exposures to $R_M R_F$, R_{SMB}, and especially R_{RMW} and R_{CMA}. Authors suggested that four factor model (without R_{HML} factor) can be used as well as five factor model for estimation of abnormal returns. As another alternative, R_{HML} can be substituted by R_{HMLO} (orthogonal HML) as the sum of the intercept and residual from the regression of R_{HML} on $R_M R_F$, R_{SMB}, R_{RMW}, and R_{CMA}.

Overall, Fama and French (2015) claim that the five-factor model is superior to the three-factor model (Fama and French 1993) although FF5 model fails the GRS test rejecting the null hypothesis that the market model pricing errors are jointly zero. According to the results, the new five-factor model explains between 71% and 94% of the cross-section variance of expected returns for the Size, B/M, OP, and Investment portfolios. However, it was noted that the model is unable to describe average returns of the small stocks of firms that invest a lot despite low profitability.

The next study of Fama and French (FF, 2016) tested the five-factor model on the range of anomalies which cannot be captured by FF3 model and were previously discussed in the literature; such as accruals, net share issues, momentum and volatility. The results showed that in contrast to the three-factor model, five-factor model's positive exposures to *RMW* and *CMA* (typical of profitable firms that invest conservatively) are able to explain the high average returns associated with low β, share repurchases, and low volatility. Contrariwise, negative FF5 exposures to *RMW* and *CMA*, (typical of less profitable firms that invest aggressively) capture the low average returns associated with high β, large share issues, and highly volatile returns. However, it was explained that the portfolios that are in the smaller *size* quintiles (microcaps) and in the highest quintiles of share issues and volatility cause serious problems when tested in respect to net share issues and volatility anomalies. Moreover, similarly to the three-factor model, accruals are still the main problem which remains unexplained by the five-factor model. The model also showed poor performance for portfolios formed on momentum. Authors emphasised that adding momentum factor is beneficial to the five-factor model and improves its explanatory power. Six-factor model (with momentum included, MOM-factor) performed best on the GRS test. Authors claim that models with MOM factor, including Carhart's (1997) four-factor model, perform almost as well as the six-factor model. Nevertheless, the six-factor model leaves lots of momentum in microcap returns unexplained.

Fama and French (2017) conducted further FF5 tests using international data from four regions: North America, Europe, Japan and Asia Pasific. The results showed that average stock returns for all regions excluding Japan increase with the book- to-market ratio and profitability and are negatively related to investment. Global FF3 and FF5 models were considered poor performing in tests on regional portfolios. For Japan, a strong positive relation between B/M and average returns is the only pattern captured by the local version of FF3 model. The factor spanning tests conducted for the period from 1990-2015 revealed that investment factor (CMA) is redundant for Europe and Japan as it adds little to the description of average returns. The study also confirmed the issues raised in FF (2015) that portfolios of small stocks whose returns behave like those of firms that invest a lot despite low profitability have low average returns and cause problems to FF5 in many different sorts.

Among other evidence on the Fama and French five-factor model performance we can mention Lin (2017) who tested the ability of the five factor model to describe average returns in the Chinese equity market over the period 1997-2015. Similarly to Fama and French (2017) they found FF5 investment factor to be redundant. In contrast to Fama and French (2015) both value and profitability factors were considered important, however, based on the results profitability has higher explanatory power than HML factor. Nichol and Dowling (2014) compared the performance of the Fama and French model with a three factor model proposed by Chen et al., (2011) consisting of the market factor, an investment factor, and a return-on-equity factor. The results suggest that investment factors appear not to be effective in the UK context, and that FF5 provides marginal improvements to the widely used FF3 model with its profitability factor offering the most potential. On the opposite side Chiah et al. (2016) compared the performance of the FF3 vis a vis FF5 model in pricing Australian equities and provided results that FF5 outperforms FF3 and is able to explain better asset pricing anomalies.

Similarly, the superiority of the FF5 in explaining the returns of anomaly portfolios was stated in Zaremba and Czapkiewicz (2017), who performed a comparative analysis of factor pricing models for Eastern European markets (Czech Republic, Hungary, Poland, Russia, and Turkey). Huyhn (2017) tested the ability of FF3 and FF5 models to explain anomalies in Australia with focus on the spread return to long-short trading strategies. The results showed significant spread returns for 16 out of 19 anomalies examined. This study confirmed that the number of anomalies that remain decreased when FF5 applied; however, it stated that the findings provide cautious support that the new factors RMW and CMA have a role to play. The work emphasised that both FF3 and FF5 models failed GRS test and concluded that the search for the most accurate asset pricing model continues. Similai (2016) asserted that FF5 can provide a parsimonious description of average returns of accrual-sorted portfolios. Ball et al., (2016) posited that FF3 and FF5 does not explain accrual anomaly and suggested to use cash-based operating profitability (a measure that excludes accruals), which is better in explaining the cross section of expected returns than gross profitability, operating profitability, and net income, all of which include accruals. According to the results, cash-based operating profitability explains expected returns as far as ten years ahead.

2.3.4. Recent evidence criticising Fama-French-Carhart portfolio/factor construction

The recent paper of Cremes, Petajisto and Zitzewitz (2012) raises the question on the validity and accuracy of the Fama and French three-factor and the Carhart four-factor model. Empirical evidence shows that the standard benchmark models produce economically and statistically significant non-zero alphas even for passive benchmark indices. Thus, when S&P 500 index performance is regressed on the Carhart model for the sample period from 1980 to 2005 the results reveal an annual alpha of 0.82% (t=2.78), with a similar result for the Russel 2000 index, an annual apha of –2.41% (t = –3.21). Therefore, a passive portfolio that is long S&P 500 Growth and short Russell 2000 Growth has a surprising annual alpha of 5.23% (t = 4.23), in other words these outcomes show that a fund/portfolio manager that simply replicates passive benchmarks can be classified as skilful, whereas in fact exhibits a median performance.

According to Cremes, Petajisto and Zitzewitz (2012) the causes of the biases are as follows: first, disproportionate weight on small value stocks. The Fama and French equal-weighted SMB factor construction underweights small value stocks in the benchmark for large-cap portfolios and overweights small value stocks in the benchmark for small-cap portfolios. Therefore, it contributes to a positive alpha for large stocks and a negative alpha for small stocks. Second, value-weighted excess return obtained from CRSP which is used as market factor comprises not only U.S. common stocks, but also non-U.S. firms, closed-end funds, real estate investment trusts (REITs), and other securities such as shares of beneficial interest (SBIs). Aforementioned assets massively underperformed U.S. common stocks from 1980 to 2005, giving an annual Carhart alpha of -4.01% (t = 2.67). Hence, the CRSP market index underperforms U.S. common stocks by about 23 basis points per year. So that gives a positive alpha for indices that are comprised of US common stocks, for instance the S&P 500. Last, annual changes to the indexes (reconstitution effect) also add to negative index alphas, especially for small-capitalization indices.

The paper suggested a modified Carhart version, and proposed index models that reduce passive alphas significantly and produce less out-of-sample tracking error volatility when used to explain actively managed mutual fund returns. In addition authors suggested a seven-factor model which captures the relative performance of midcaps and allows the value-growth effect to differ for large, midcap, and small cap stocks. The study shows that the standard Fama and

French and Carhart models can significantly bias inferences about manager performance. Thus, according to the unadjusted Carhart model small-cap managers underperform large-cap managers by −2.13 percentage points per year, whereas this result is fully reversed when the performance is estimated with any the altered or index factor model proposed.

It is not a first study questioning Fama-French-Carhart portfolio construction. Chan, Dimmock, and Lakonishok (2009) also report a negative and statistically significant alpha for the Russell 2000 Growth index (their Table 8). Costa and Jakob (2006) document significant non-zero alphas and significant factor loadings on the momentum factor for a large set of stock market indexes and state that the benchmark model biases can affect conclusions about managerial performance. The study of Huij and Verbeek (2009) argue that factor proxies systematically bias the multifactor performance estimates of mutual funds and claim that it comes from miscalculating the factor premiums which are either over- or underestimated.

2.3.5 Fund reference benchmark. Evidence of fund performance manipulation and methodologies proposed to eliminate biases.

It is common for investors to make judgements on mutual fund performance by comparing the excess return of a fund to a return of a passive benchmark with similar risk characteristics. However, recent literature claims that it might be more accurate to use fund reference benchmark instead. For instance, Angelidis, Giamouridis and Tessaromatis (2013) argue that manager skill/performance should be measured relative to their self-reported benchmark as using a passive portfolio with the same risk characteristics instead may misstate the performance. In the US SEC regulations require mutual fund companies to disclose their performance relative to a passive benchmark. This benchmark (index) is often referred as a fund prospectus benchmark and is commonly used for performance evaluation purposes among academics and practitioners.

However, there is a new flow of literature which raises concerns on the content of information disclosed, in particular, on the accuracy of the fund prospectus benchmarks. Some authors claim that mutual funds do not consistently follow the investment style described or tend to rotate their portfolios in order to manipulate the results by possible risk shifts. Sensoy (2009) find evidence that self-designed benchmarks are constantly mismatched by mutual funds. Thus, results show that 31.2 % of funds analysed specify a benchmark index that is misspecified versus S&P or Russell size and value/growth-based benchmarks (“corrected” benchmarks, which better match funds’ size and value/growth characteristics, and are more correlated with funds' returns.). The work claims that it is not incidental. Some funds tend to use misleading benchmarks, so that their performance looks better and generates more subscriptions. Thus, evidence shows that value funds are more likely to have self-designated benchmarks that are mismatched on value/growth, small-cap funds tend to have prospectus benchmarks mismatched on size. Based on the fact that misleading is more common among large and high-fee funds, paper concludes that benchmarks are mismatched for strategic reasons.

Cremers and Petajisto (2009) provide evidence that mutual funds typically have a high proportion of holdings that differ from those of the fund’s theoretically correct benchmark index. Substantial exposures to size and value/growth factors in returns that are not captured by their benchmarks were also discussed in Elton, Gruber, and Blake (2003).

Kim, Shukla and Tomas (2000) examine of how well mutual funds’ stated objectives conform to their attributes-based objectives. The study compares information disclosed by funds to the funds’ attributes, grouped by characteristics, investment style, and risk/return, obtained from Morningstar database. The findings show that the stated objectives of more than half of the 1043 funds analysed differ from their attributes-based objectives, and over one third of the funds are severely misclassified. Nonetheless, authors state that it is not always that all funds deviate into higher risk objectives. In fact, based on the results some funds tend to diverge into lower risk objectives. Therefore, the research concludes that this tendency cannot be explained by gaming behaviour.

DiBartolomeo and Witkowski (1997) examine monthly returns for 748 load and no-load open-end funds and show that return patterns of 40 percent of funds analysed deviate from the benchmark declared in the prospectus with 9 percent of funds being seriously misclassified, two or more risk tiers away from their declared categories. The reclassification matrix displays that observed misclassification took place in both directions, upwards, into more aggressive categories, and downwards (those funds appear to be less aggressive than their group peers).

Bams, Otten, and Ramezanifar (2016) analyse a sample of 1,866 US equity funds over the 2003-2015 period and provide evidence that 14% of funds are significantly misclassified based on long term style analysis. The performance analysis conducted based on the Carhart model shows that in the long run misclassified funds significantly underperform well-classified funds by 0.92% per year on a style-adjusted return basis and by 1.18% per year on a net return basis, respectively. The results reveal that misclassified funds tend to be younger, smaller in size and charge higher expense ratios.

Among others Castellanos and Alonso (2005) provide evidence of misclassification in Spanish mutual funds. Huang, Sialm and Zhang (2011) show, that mutual funds change their total risk exposure substantially over time. Authors claim that it might be done for strategical reasons: in order to increase the expected money inflows to the funds or to manipulate their performance numbers. Similarly, Kacperczyk, Sialm, and Zheng (2008) measure the return gap, the difference between the actually generated return and the expected return on the hypothetical portfolio of previously disclosed fund holdings. The study documents the persistent in the long run effect of unobserved actions for both the worst and the best performers and claims that it can be used to predict their future performance. The evidence shows that the hypothetical holdings returns for the most funds have large correlations with the investor returns; hence the investment strategies do not deviate significantly from their disclosed strategies. However, there are cases for funds with relatively low correlations between holdings and investor returns. In addition, they show results consistent with Chen et al. (2004) that smaller funds and larger funds families tend to exhibit more favourable return gap.

Wermers (2012) emphasizes that portfolio "style drifts" (shifts in loadings on priced style factors or style characteristics) can be a substantial source of risk for those who invest in the funds as it is almost impossible to monitor every manager trade, and normally trades are disclosed with noise and a significant time lag. The study analyses the causes and consequences of portfolio "style drift" among U.S. equity mutual funds and proposes a new "holdings-based style drift measures".

Brown, Harlow, and Starks (1996) claim that since the compensation of the money managers is directly linked with their performance fund managers in the top performing funds may have different attitude to risk versus others who perform badly. They provide evidence that managers with the worst performance by the middle of the year take on excessive risk to improve their performance numbers. Similarly, Chevalier and Ellison (1997) emphasise the link between the flow-performance relationship and managerial fees and state that mutual funds tend to alter the riskiness of their portfolios at the end of the year to increase the inflow of investments.

Chan, Chen, Lakonishok (2002) provide evidence of subsequent shifts in style of mutual funds with poor past performance. The results show that such style drifts are particular notable in the case of value funds which have experienced poor past performance (versus growth funds with poor returns). Style-switching behaviour of mutual fund managers was also discussed in Frijns, Gilbert and Zwinkels (2016). The results indicated the evidence of “twin-style” switching within the value-growth dimension and within the small-large dimension. Significant switching was documented for close to 53% of the funds analysed, where the most significant switching was observed for the the Mid-Cap Value Equity funds (68%) and the least significance reported for the Large-Cap Growth Equity funds (44%).

Goetzmann et al. (2007) affirms that active managers prone to change risk levels to manipulate their investment record and explain how by simple dynamic manipulation strategies possible to change risk levels and influence results. The study concludes that by changing risk levels active portfolio managers may jeopardise the overall performance of the fund. Jennifer, Sialm, and Zhang (2011) provide evidence that funds that tend to shift risks perform worse than funds that maintain stable risk levels. The results indicate potentially severe consequences for funds that increase their risk levels. As an example they apply the Carhart model and show that funds that are in the highest risk shifting decile exhibit an abnormal return of -29 basis points per month. They suggest that risk –shifting can be viewed as an indicator of inferior ability, poor managerial skills and agency issues.

Thus, for more accurate fund performance assessment Cremers and Petajisto (2009) propose to use Active Share measure which represents a fraction of the portfolio holdings that is different from the portfolio’s benchmark index. They claim that in combination with tracking error Active Share helps to determine the actual type of active fund management. The results show that funds with high Active Share significantly outperform their benchmarks, both before and after expenses, and show strong performance persistence. Chan et al. (2009) proposed to estimate performance relative to characteristic-matched benchmarks (constructed based on size-conditional book-to-market sorts, quarterly size-conditional book-to-market sorts, size-conditional composite value/growth indicator approach) and the Russell style indexes. For the latter model a corresponding Russell index was assigned to each active portfolio according to its style, where styles were obtained from the reports provided by money managers. The results show that inferences about performance are sensitive to the benchmarking methodology.

The recent study of Hunter at al., (2014) proposed a novel methodology which provides a solution to the issue of non-zero benchmark alphas generated by standard three and four-factor models and helps to eliminates biases caused by inaccurate fund self-reported benchmarks. Thus, the work explains that prior making a choice on the best mutual fund investors identify the group of funds which according to the investment style/objectives/risk undertaken suits them best. Then, they select the fund within the group with the best past performance. Therefore, authors do not target to estimate the alpha of a fund, they explain that the most beneficial for investors is to understand the performance of the fund relative to peers in the reference group. For the most accurate analysis the true manager skills/performance should exclude the commonalities in fund strategies within a peer group, therefore should account for information on fund returns and investment objectives. The study claims that instead of trying to augment commonly used factor models by adding numerous exogenous factors with many complex strategies that could be implemented within a peer group, it is easier to build an additional benchmark, so-called active peer benchmark (APB), based on the return on the group of funds to which a given fund belongs. Hence, only one factor added to the standard benchmark models will account for peer group commonalties in idiosyncratic risk-taking and will allow estimating unique manager skills that are uncorrelated with the manager's active peer group's average skills.

To construct Active Peer Benchmark authors use the return of the equal-weighted active peer group that a fund belongs and calculate the Active Peer Benchmark's Carhart alpha and Carhart error term. The sum of both, APB alpha and APB error term, represents the additional APB factor (the fifth-factor) in the augmented Carhart model proposed. Thereby, if a fund manager has skills that are above common strategies/practices used within the reference group, the APB-adjusted alpha in the new APB adjusted model will be positive and significant. Based on outcomes it is easy to rank the funds within the reference group and identify subgroups of funds with the top skills. In other words, active peer benchmark can be considered as a zero skill asset and can be viewed as a passive benchmark for a fund. The tests of the model shows that APB adjusted Carhart version significantly reduce average time series correlation of residuals between individual funds within a peer group. Moreover, based on the results, adding commonalities in fund strategies to the standard benchmark model factors result in greater explanatory power of the APB adjusted Carhart model. Robustness tests revealed that APB-adjusted model outperforms the ranking of the model proposed by Cremers and Petajisto (2009).

2.4. Existing evidence on mutual fund performance

Prevailing academic evidence on fund performance shows that after controlling for fees active portfolio managers do not beat their benchmarks. Sharpe (1991) in the study “Arithmetic of Active Management” explains that the performance of the index equals the weighted-average return of both active and passive investors before investment expenses. Since by definition, active managers bear greater costs, their after-cost return must be lower than that from passive management. Hence, he states that active management is zero sum game leading to the conclusion that active fund managers cannot beat the returns generated from passive investment strategies. In support of this statement Cruber (1996) show that the average actively managed funds underperform their benchmarks where half of the underperformance was driven by expense ratio. They share the same view and claim that investors would be better off by investing in passively managed index funds which provide almost the same or better performance at lower costs. According to Carhart (1997) mutual funds do not earn considerably high alpha, the results are relative and after deduction of transaction costs and expenses the majority of funds underperform. Only top mutual funds generate higher returns that can cover expenses. The bottom-decile funds underperform by about twice their reported investment costs. French (2008) asserts that the performance of active funds excluding expenses, trading costs and fees is a negative sum game. Based on reasonable assumptions, the work explains that if investors divert the resources they spend trying to beat the market into passive market portfolio, they would increase their average annual return by 67 basis points over the 1980-2006 period. Wermers (2000) examine U.S. domestic equity fund over the 1975–1994 period and provide evidence that, on average, funds tend to underperform its overall market, size, book-to-market, and momentum benchmarks by 1.2 per cent per year. Barras, Scaillet, and Wermers (2010) document that 75 per cent of mutual funds can generate excess performance and show some stock-picking ability, however it is just enough to cover their fees. Other 24 per cent funds underperformed with significant negative alpha and only 0.6 per cent funds outperformed with the following results being statistically indistinguishable from zero. The study shows a dramatic decline in the proportion of skilled funds from 14.4 per cent in early 1990 to 0.6 percent in late 2006 with the inverse tendency for unskilled funds where the highest proportion of unskilled funds observed among larger and older funds. Similar findings were described by Chalmers et al. (2001) who verify that active funds do not produce high enough gross returns to cover the average trading costs. They conclude that in the long perspective portfolio revision damage shareholders’ value. Becker, Ferson and Schill (1999) apply conditional market-timing models and find no evidence on the mutual funds market timing ability. Jiang (2003) use a large sample of actively managed domestic equity funds that have different benchmark indices and confirm that the results do not show superior timing abilities among funds.

Kacperczyk, Sialm and Zheng (2005) investigate whether some fund managers create value by concentrating their portfolios in industries where they have informational advantages. The results indicate that more concentrated funds perform better after adjusting for risk and style differences and yield an average abnormal return of 1.58% per year before deducting expenses and 0.33% per year after fees. Chan, Chen and Lakonishok (2002) show evidence that funds tend to cluster around the benchmark with only few of them that deviate from the index; they typically allocate their portfolios into growth and past winning stocks. Funds with poor past performance tend to rotate their portfolios with higher frequency. The results show that money managers do not possess skills to time movements in the style factors. Nevertheless, the work provides evidence that growth funds on average do better than value funds.

However, there is a growing strand of academic literature which claims that fund outperformance can be observed during the periods of booms/recessions and different phases of the business cycle. Moskowitz (2000) demonstrate that actively managed funds generate an additional 6%/year during recessions versus expansions. Kacperczyk, Nieuwerburgh and Veldkamp (2014) claim that managerial stock picking and timing skills should be examined in the periods of booms and recessions. They apply a new measure of managerial ability and show evidence that top funds are able to outperform other funds and passive benchmarks and generate the CAPM, three-factor and four-factor alphas of 50 to 80 basis points per year in excess of the other funds. They examine the funds' characteristics and show that the top funds are represented by younger funds, with higher expenses and less wealth under management, tend to exhibit higher portfolio turnover and receive higher inflows of new assets to manage; their betas typically deviate more from the peers and the portfolios are constructed with fewer stocks and have higher stock- and industry-level dispersion. In recession the top funds significantly increase cash holdings, and rotate their portfolio allocation from high to low market betas by increasing their portfolio weights in defensive industries , whereas, in boom they shift their portfolios into high betas sectors and cyclical industries. The results on the performance persistence show that superior performance can be observed for the next year, statistically significant for up to six months.

Lucas, Dijk and Kloek (2002) use statistical time-series model and macroeconomic regression models and find that style investing based on macroeconomic predictors and controlled for business cycles provides significant risk-adjusted excess returns. Petajisto (2013) using Active Share and tracking error analyses the performance of all-equity mutual funds during the 2008-2009 financial crisis and shows that the active stock pickers beat their indices by about 1%, whereas the closet indexers underperformed. The paper also documented negative relationship between the size of mutual funds and their performance.

Other researchers claim that in order to observe excess returns mutual fund performance should be analysed during short time periods. For instance Bollen and Busse (2005) state that stock selection and market timing skills persist during short term and disappear when funds are evaluated over longer periods. Zheng (1999) provide evidence of short-term positive excess returns for funds with positive new money flow. However, evidence does not stand when it is tested for longer periods. The study affirms that the results do not confirm that active portfolio managers on average can beat the market; nevertheless, it shows that small funds do. The results are robust for conditional and unconditional performance measures.

Herrmann and Scholz (2013) examine 520 hybrid mutual funds over the period 1998-2009 and show that, on average, hybrid mutual funds do not outperform their benchmarks. They state that hybrid mutual funds exhibit short-term persistence in in-quarter abnormal performance; however, it is not clear whether they possess abilities to successfully shift style exposures on a quarterly basis. Moreover, Davis (2001) used Fama and French methodology to construct style-based portfolios in order to analyse the relationship between manager style and equity fund performance. The results suggest that no investment style in the study earned excessive returns during the period analysed. Only short-run performance that did not stand beyond the year was observed for the best-performing firms. At the same time they document negative abnormal return of about 2.75 percent for the value funds.

Shukla (2004) analyse whether active portfolio management create value to the shareholders and estimate active fund performance from interim portfolio revisions by taking the difference between the actual return on portfolios and the buy-and-hold returns. Results show that for a 1 months holding period almost 50 per cent managers were able to generate excess gross returns, some of them even up to 3 per cent per month. However, due to the fact that funds that deliver higher returns require higher expense ratio, excess return was wiped out after controlling for trading costs. The analysis for longer holding periods of up to 6 months provided even worse results with negative median and average return.

Karoui and Meier (2009) examined 828 newly launched US equity mutual funds for the period from 1991 to 2005 which, on average, generate higher excess and abnormal returns with the risk adjusted performance which is superior to existing funds. The work confirm the short-term persistence among top-performing funds, however, it shows that this performance does not hold more than two subsequent periods. It provides the evidence of a significant drop in performance, with a rapid fall from the top to the bottom decile, for a substantial fraction of funds. To identify the possible reasons the work analysed portfolio characteristics of the funds and explained that the portfolios are typically concentrated in smaller and less liquid stocks, less diversified in terms of number of stocks and industry concentration and, as a consequence, returns of the funds demonstrated higher ratios of unsystematic to total risk.

2.4.1. Empirical evidence in the UK

In comparison with the U.S. studies in mutual fund performance, there have been comparatively few studies examined the ex-post performance of mutual funds (unit trusts) in the UK. Consistent with the U.S. evidence Cuthbertson, Nitzsche and O'Sullivan (2008) continue the evidence applying cross-section bootstrap methodology and report stock picking ability among a relatively small number of top performing UK equity mutual funds suggesting that UK equity investors will be better off holding index/tracker funds. The work of Cuthbertson, Nitzsche and O’Sullivan (2012) based on false discovery rate approach show that only around 3.7% of all funds truly outperform their benchmarks versus 22% of funds which truly underperform their benchmarks. Blake and Timmermann (1998) examined UK open-end mutual funds over a 23-year period and find evidence of underperformance on a risk-adjusted basis by the average fund manager which is in line with the U.S. evidence in mutual fund performance. Moreover, the results of the paper point out on weekly outperformance among mutual funds during their first year of existence. Another work of Cuthbertson, Nitzsche and O'Sullivan (2010) use a recent nonparametric methodology and tests the market timing ability of UK mutual funds. The outcomes show that relatively small number of funds (around 1%) show positive market timing ability while around 19% of funds exhibit negative timing. Overall, results demonstrate that on average funds miss-time the market. Byrne, Fletcher and Ntozi (2006) provide similar results highlighting no evidence of superior market timing performance and find that UK unit trusts act like benchmark investors. Quigley and Sinquefield (2000) use the three-factor model to estimate the performance of UK unit trusts that concentrate their investments in UK equities. The results are consistent with the evidence and show that money managers underperform the market with the worst performance verified for small-company trusts. Evidence on performance persistence shows that only bad performance does persist.

Interesting evidence provide Fletcher and Forbes (2002) who investigate performance persistence in UK mutual funds for the 1982-1996 time period by applying different models: CAPM, APT and the four-factor Carhart (1997) model. The findings lead to different results. The work verifies significant performance persistence for the portfolios of unit trusts, formed on the basis of prior year excess return, when returns are compared with first two models; but persistence disappears when performance is evaluated using the four-factor model. However, conditional performance measure of Ferson and Schadt (1996) reversed this result with even stronger evidence of statistical significance.

Chapter 3
First empirical essay

Abstract

In this study we re-visit the performance of 887 active UK equity mutual funds using a new approach proposed by Angelidis, Giamouridis and Tessaromatis (2013). The authors argue that mutual funds stock selection is driven by the benchmark index, so if the benchmark generates alpha, there will be a bias in interpretation of manager's stock picking ability. In their model, alpha of a fund is adjusted by benchmark's alpha. By applying this method, we eliminate bias inflicted by the persistently negative alphas of FTSE 100 index in the period 1992-2013. We find that adjusted Fama-French and Carhat alphas of UK equity mutual funds are higher than those implied by the standard three and four factor models and overall positive, contrary to most of the existing literature on UK fund performance. This result is consistent across funds' investment styles and robust to use of FTSE Small Cap as benchmark for a subsample of small cap funds.

Keywords: Fama-French, Carhart, adjusted alphas, UK equity funds performance
JEL classification: G11, G12, G23

3.1. Introduction and Literature Review

In this paper we re-visit the question of performance of active UK equity mutual funds by modifying the Fama and French (1993) three-factor (FF3 hereafter) and the Carhart (1997) four-factor models using Angelidis, Giamouridis and Tessaromatis (2013) approach. The FF3 and Carhart models are widely accepted, standard methods of estimation of abnormal returns and portfolio manager's security selection ability (alpha) by many researchers, investors and investment practitioners internationally. Angelidis et al. (2013) argue that security selection in a fund is largely driven by the composition of a selected benchmark; so if the benchmark itself generates significant out/underperformance in the standard performance evaluation models, then investors' interpretation of manager's stock picking ability is biased. To correct for this bias, authors alter the left-hand side of the Carhart (1997) model by replacing excess return of a fund (relative to the risk free rate) with benchmark-adjusted return. Their modification produces a new fund alpha adjusted for the alpha embedded in the benchmark. Such a new alpha therefore represents managers' 'true' stock-picking ability. Angelidis et al. (2013) test the model on a sample of US equity mutual funds and report that benchmark-adjusted alphas are less negative and less statistically significant than the Carhart ones. We believe that this method provides a useful novel insight into performance measurement that is of interest to academics and, in particular, investment professionals. We contribute to the literature by being the first study to re-visit UK equity mutual fund performance utilising this new methodology.

A significant strand of recent academic literature criticises FF3 and Carhart models with specific emphasis on the fact that factor misspecification leads to presence of non-zero alpha in passive indices used as benchmarks in performance measurement. For instance, Chan, Dimmock, and Lakonishok (2009) report a negative and statistically significant alpha for the Russell 2000 Growth index. Cremers, Petajisto and Zitzewitz (2012) reveal an annual Carhart alpha in the S&P 500 index of 0.82% (t=2.78) and in the Russel 2000 that of –2.41% (t = –3.21) for the sample period from 1980 to 2005. Such positive (negative) index alphas would create an upward (downward) bias in a performance of funds benchmarking against those indices.

One of the explanations of significant index alphas offered in literature is the error in the construction of the risk factors FF3 and Carhart models specify; namely: the market risk premium, the size factor (SMB, defined as the return of the small capitalisation minus the return of the large capitalisation portfolio), the style factor (HML, defined as the difference in returns of high vs. low book-to-market ratio stocks. i.e. value vs growth stocks) and the momentum factor (return difference between past winners and past losers portfolio). Cremers et al. (2012) suggest several causes of these errors: first, the FF3 model overweights stocks in the small value portfolio, which outperformed during the specified time period, exaggerating the return on the SMB factor; second, value-weighted excess return obtained from CRSP includes non-U.S. shares, which underperformed U.S. common stocks during the sample period; third, annual changes to the indexes contribute to negative index alphas especially for small-capitalization indices. The authors propose reconstruction of factors to obtain modified Fama-French-Carhart models that reduce passive index alphas significantly and produce less out-of-sample tracking error volatility when used to explain actively managed mutual fund returns. Similarly, Huij and Verbeek (2009) argue that factor proxies systematically bias the performance estimates of mutual funds caused from miscalculating the factor premiums which are either over- or underestimated. Costa and Jakob (2006) document significant non-zero alphas and significant factor loadings on the momentum factor in the Carhart model for a large set of stock market indexes. Recent Fama and French (2012) study confirms that there is a concern with factor portfolios formed on size and momentum in the FF3 model. They examine the size, value, and momentum in individual stocks returns across four regions (North America, Japan, Europe and Asia-Pacific) to test whether the value and momentum patterns in international returns are captured by FF3 and Carhart models. The results show consistent risk premia across markets.

Given the evidence from these studies, some adjustments to the existing FF3 and Carhart models are of essence for the improved performance measurement. One such adjustment is related to incorporating the fund's benchmark returns in the models, as highlighted by Cremers, Petajisto and Zitzewitz (2012). Hsu, Kalesnik and Myers (2010) propose a dynamic allocation attribution methodology based on the traditional Brinson attribution. It includes the adjustment for static and dynamic factor allocation and authors state that "normal portfolio" which represents a manager's preferred allocation can be used as a benchmark when no explicit benchmark exists. Further, Angelidis, Giamouridis and Tessaromatis (2013) adjust the mutual fund returns for the returns of the fund's self- reported benchmark. They argue that a mutual fund performance should be measured relative to its self-designated benchmark and the use of market implicit benchmark rather than self-designated benchmarks biases the current academic performance evaluation practices. Nevertheless, Angelidis et.al (2013) state that their approach can utilise any benchmark a fund wishes to measure their performance against. Note that the choice of self-reported benchmark by funds is not always clear. For example, Sensoy (2009) finds evidence that self-designated benchmarks are persistently mismatched by mutual funds, which may be explained by the funds' strategic incentives to improve inflows. The paper stresses the need for the development and dissemination of measures of mutual fund performance that are both well-grounded in economic theory and not subject to gaming.

Let us review now what is known so far about UK fund performance. Relative to the large number of U.S. studies[2], there has been comparatively fewer studies examining the ex-post performance of mutual funds in the UK. The vast majority of existing UK studies utilise standard unconditional CAPM, FF3 and Carhart models to estimate fund alphas. UK studies corroborate findings from the US, providing stronger evidence in support of fund underperformance than fund outperformance. For instance, Cuthbertson, Nitzsche and O'Sullivan (2008) apply cross-section bootstrap methodology and report stock picking ability among a relatively small number of top performing UK equity mutual funds suggesting that UK equity investors will be better off holding index/tracker funds. Further work by Cuthbertson, Nitzsche and O'Sullivan (2012) based on false discovery rate approach show that only around 3.7% of all funds truly outperform their benchmarks versus 22% of funds which truly underperform their benchmarks. Earlier studies such as Blake and Timmermann (1998) assess UK open-end mutual funds and find evidence of underperformance on a risk-adjusted basis by the average fund manager. On a positive note, they point at the weekly fund outperformance within the first year of inception.

Since the FF3 and Carhart model alphas are commonly used as measures of performance in the range of studies discussed here, and the recent literature points at biases in these measures, we believe that re-evaluation of UK mutual fund performance based on *adjusted* factor model is needed. In this study, we contribute to the UK mutual fund performance literature by applying Angelidis et al. (2013) methodology to re-examine the performance of active UK equity mutual funds. We do not claim that Angelidis et al. (2013) is the best model for adjusting FF3 and Carhart alphas and it is not the purpose of the present paper to determine which one is. However, it is a model that not only has academic rationale, but it may also resonate well with

[2] See for instance: Pastor and Stambaugh (2002), Wermers (2000), Daniel et al. (1997), Carhart (1997), Grinblatt et al. (1995) among others.

practitioners as a) it is less computationally intense than some of the models that require reconstruction of risk factors (e.g. Cremers et al., 2012) and b) it transforms the left hand side of the FF3/Carhart model into excess return of the fund relative to the benchmark; which is core to determining a fund's tracking error – a primary risk matric for investment professionals. To the best of our knowledge, this paper is the first study of active UK equity mutual fund performance using Angelidis et al. (2013) proposed adjustment to the standard FF3 and Carhart models. Such a new take on performance will help investors shed a better light on the choice between active and index tracker funds and revise the previous work on UK fund performance. Therefore, the key question in this paper is whether UK equity fund performance is underestimated by traditional models and whether investors on average actually generate better alphas than existing evidence might suggest. The reader should note that while alpha of a fund may be biased, a fund's ranking may not be; nevertheless, the research into individual fund rankings before and after alpha adjustment is not the focus of this paper.

Our data set includes 887 active UK equity mutual funds for the period January 1992 to October 2013 and FTSE 100 index as the benchmark for the UK focused funds. We identify a significant *negative* FF3 alpha of -1.12% and Carhart alpha of -1.13% of the FTSE 100 index, implying a downward bias in fund performance. Further, we document that this bias is stronger in bear markets. During our period of analysis, the standard FF3 alpha for our sample of funds is only 14bps per year and not strongly significant. By applying the Angelidis et al. (2013) method, the adjusted FF3 alpha of 887 funds increases tenfold to 144bps, significant at 1% level. Similar strong improvement in alphas is confirmed across the bull and particularly the bear sub-sample periods. We also examine if good (or bad) performance is particularly related to an investment style of a fund. Splitting the funds into Morningstar style box categories, we document that performance was biased downward across all style categories when standard FF3 and Carhart models are used. After model alteration, we report that Small/Value and Small/Growth categories are generating positive adjusted alphas in four out of five sub-sample periods, making them the most successful segments of the market in the period analysed. We test the robustness of these results by replicating the analysis for a sub-sample of small capitalisation funds using FTSE Small Cap as a benchmark. We confirm all the previous findings and report even stronger significance of both adjusted alphas and the differences between adjusted and standard alphas. Overall, our study shows that standard FF3 and Carhart models amplify the underperformance of mutual funds reported in previous literature. Our computation of adjusted alphas proves that UK equity fund performance is better than initially documented and significantly positive.

The remainder of the paper is organised as follows: Section 2 describes our data, Section 3 presents the methodology, Section 4 lays out the main results, Section 5 presents robustness check and Section 6 concludes the paper.

3.2. Data and Methodology

3.2.1 Data description

The data set comprises of 887 actively managed equity mutual funds with UK investment focus. The net monthly returns of mutual funds are from Morningstar, inclusive of dividends. There is no survivorship bias in the sample. We use FTSE 100 index as a benchmark for measuring performance of our funds. This index represents 80% of UK market capitalisation and is commonly regarded as a proxy of the UK market performance. While our funds follow various investment styles (from Morningstar style box, as discussed in Section 4.3), indices covering combinations of styles such as medium/value, small/growth etc. are not available in the UK and mainstream UK funds still resort to a general market index as benchmark[3]. Therefore, we choose an index commonly used to represent UK market trends – the FTSE 100. The returns of the FTSE 100 index (inclusive of dividends) are from Datastream. We provide a robustness check with FTSE Small Cap index as a benchmark for funds in the small capitalisation category and provide a short discussion on use of other style benchmarks in Section 5. The monthly FF3 and Carhart factors for the UK, as well as the UK risk free rate are defined as in Gregory, Tharyan and Christides (2013) and obtained from University of Exeter, Xfi Centre for Finance and Investment website[4]. The period of analysis spans from January 1992 to October 2013. We split the sample into five bull and bear sub-periods, as follows: January 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull markets); January 2000 to December 2002 and January 2008 to December 2009 (bear markets)[5].

The following Table 1 provides descriptive statistics of the funds comprised in the sample. The main requirements for the funds to be included are to have at least 36 months of continuous

[3] http://www.morningstar.co.uk/uk/news/58752/understanding-benchmarks.aspx
[4] http://business-school.exeter.ac.uk/research/areas/centres/xfi/research/famafrench/files/
[5] The FTSE 100 annualized return for the five periods analysed are the 11.04%, 9.22% and 9.25% (bull market periods) and -19.07% and -6.70% (bear market periods). We consider the dot.com bubble burst and recent financial crisis as bear periods.

observations for the total period and a minimum of 30 months of continuous returns within each rolling period. The number of funds and monthly observations are presented for each rolling window. We also provide separate figures (funds/monthly observations) for bull and bear market periods. As it can be seen from the Table the number of funds and consequently monthly observations increases towards the last years. In total, the sample is comprised of 887 funds with 123,768 monthly observations (no overlapping).

Table 1: Funds with more than 36 monthly observations

Table reports the number of funds and monthly observations for each of the 36 months rolling windows. The minimum data requirement is for funds to have at least 36 months of continuous observations and no less than 30 months of continuous returns within each rolling period. The #Funds represents the number of (non-unique) funds with available data in each period.

Period	*# Funds*	*# Monthly Observations*	*Period*	*# Funds*	*# Monthly Observations*
199201:199412	239	7,908	*200201:200412*	595	18,928
199301:199512	270	8,526	*200301:200512*	631	20,540
199401:199612	287	9,222	*200401:200612*	715	22,315
199501:199712	306	9,990	*200501:200712*	760	23,884
199601:199812	336	10,762	*200601:200812*	789	24,992
199701:199912	385	11,762	*200701:200912*	801	25,204
199801:200012	412	13,007	*200801:201012*	776	24,548
199901:200112	454	14,272	*200901:201112*	766	23,525
200001:200212	502	15,690	*201001:201212*	727	22,404
200101:200312	550	17,204	*201101-201310*	691	20,363
Bull Market					
199201:199912	385	26,107			
200301:200712	760	37,060			
201001: 201310	735	28,140			
Total	**881**	***91,307***			
Bear Market					
200001:200212	504	15,690			
200801:200912	765	16,771			
Total	**809**	**32,461**			
Overall: No overlapping	**887**	**123,768**			

Table 2 shows summary statistics of the mutual funds' annualized excess returns. The annualised values are calculated for each rolling window and for bull and bear market periods. Mean and standard deviation of FTSE100 market risk premium is provided for a comparative purpose.

Table 2: Summary statistics

Table displays the mean, median, 75th and 25th percentile mutual fund annualized excess returns (annualised values estimated for all mutual funds in the sample). Standard deviation refers to standard deviation of average monthly return (36 months). Mean and standard deviation of FTSE100 market risk premium is provided for a comparative purpose.

Period	***Excess Return(Mutual Fund)***					***Market Risk Premium (FTE100)***	
	Mean	***p75***	***Median***	***P25***	***Stdev.***	***Mean***	***Stdev.***
199201:199412	7.570%	10.893%	8.205%	5.325%	16.401%	4.988%	15.033%
199301:199512	10.311%	13.427%	9.280%	6.709%	11.842%	8.253%	11.642%
199401:199612	6.591%	8.455%	5.077%	3.265%	10.780%	6.100%	10.360%
199501:199712	13.514%	15.962%	13.851%	11.229%	10.119%	16.613%	10.333%
199601:199812	8.724%	12.304%	10.302%	6.188%	14.318%	13.456%	13.516%
199701:199912	16.847%	18.065%	13.825%	10.752%	15.931%	14.816%	14.240%
199801:200012	6.649%	9.059%	4.828%	2.580%	17.133%	2.206%	14.728%
199901:200112	0.070%	3.798%	-2.373%	-5.448%	17.678%	-7.289%	14.072%
200001:200212	-15.079%	-12.176%	-16.571%	-19.103%	18.173%	-19.332%	16.717%
200101:200312	-3.837%	-0.816%	-7.752%	-10.448%	17.369%	-10.095%	17.287%
200201:200412	3.228%	7.043%	0.579%	-2.456%	14.099%	-1.947%	15.268%
200301:200512	15.567%	18.224%	14.269%	12.195%	11.124%	12.088%	9.962%
200401:200612	13.666%	15.945%	12.448%	10.426%	9.120%	10.489%	7.092%
200501:200712	6.425%	9.914%	7.790%	5.189%	10.350%	8.420%	8.072%
200601:200812	-11.387%	-6.910%	-9.780%	-13.557%	17.438%	-8.823%	15.309%
200701:200912	-4.422%	-2.128%	-4.773%	-8.682%	19.913%	-4.222%	18.486%
200801:201012	-0.638%	4.437%	1.347%	-2.296%	22.199%	-1.465%	20.369%
200901:201112	14.146%	18.234%	13.590%	11.094%	17.710%	11.620%	16.241%
201001:201212	9.564%	11.834%	8.739%	6.770%	14.409%	6.077%	13.388%
201101-201310	10.564%	14.453%	10.673%	8.065%	12.794%	8.008%	11.775%
Bull Market							
199201:199912	15.237%	14.909%	11.861%	10.269%	14.931%	11.660%	13.133%
200301:200712	8.341%	11.784%	9.588%	7.119%	10.807%	9.053%	9.234%
201001: 201310	13.073%	16.117%	12.512%	9.709%	14.107%	9.043%	13.223%
Bear Market							
200001:200212	-15.079%	-12.176%	-16.571%	-19.103%	18.173%	-19.332%	16.717%
200801:200912	-6.454%	-3.208%	-5.658%	-9.873%	23.942%	-7.111%	21.761%

3.2.2. Preliminary Analysis: Alpha of the FTSE 100 Index

The Fama and French (1993) three-factor model and the Carhart (1997) four-factor model have been accepted in the industry as standard models for assessing portfolio alpha-generating ability. As discussed earlier, recent literature such as Cremers et al. (2012), Chinthalapati et al., 2017 points at the presence of significant positive or negative alphas in US passive benchmark indices. We therefore start by assessing the monthly level of FF3 alpha (equation (1)) and Carhart alpha (equation (2)) over the time period *u* in the passive index commonly used as the UK market benchmark – the FTSE 100:

$$R_{FTSE100,t,u} - R_{F,t,u} = \alpha_{FTSE100} + \beta_M \left(R_{M,t,u} - R_{F,t,u}\right)_{t,u} + \beta_{SMB} SMB_{t,u} + \beta_{HML} HML_{t,u} + e_{t,u} \quad (1)$$

$$R_{FTSE100,t,u} - R_{F,t,u} = \alpha_{FTSE100} + \beta_M \left(R_{M,t,u} - R_{F,t,u}\right)_{t,u} + \beta_{SMB} SMB_{t,u} + \beta_{HML} HML_{t,u} + \beta_{WML} WML_{t,u} + e_{t,u} \quad (2)$$

where, $\alpha_{FTSE100}$ is the monthly excess return of the FTSE100 for period *u*; *t* is the frequency of the data (months) and *u* represents the estimated subperiods in months. $R_{FTSE100,t,u}$ is the total index monthly return (inclusive of dividends), $R_{Ft,u}$ is risk free rate, $R_{Mt,u}$ is the total monthly return (inclusive of dividends) of the UK equity market proxied by FTSE All Share Index as defined in Gergory et al. (2013), *SMB* and *HML* are Fama and French (1993) size (small minus big returns) and value (high minus low book-to-market returns) factors respectively, *WML* is Carhart (1997) the momentum (winner minus loser returns) factor, $e_{t,u}$ is error term.

If the performance estimation models in equations (1) and (2) are correctly specified, the FTSE 100, being a broad passive index, should not generate any (positive of negative) abnormal return. However, if it does, the performance of a mutual fund benchmarking against that index will be biased upward (if the index alpha is positive) or downward (if the index alpha is negative).

Figure 1 illustrates 3-year moving average of FTSE 100 (FF3 and Carhart) alphas over our sample period. Thus, to start we estimate FTSE 100 alphas (for both models) for the sub-period January 1992- December 1994, then continue for the sub-period January 1993-December 1995 and so on. Year 1994 on the graph represents the first sub-period, years 1995 – the second and thus continues. The alpha values are annualised and given in basis points, therefore the following formulae has been applied $[(1 + monthly\ alpha)^{12} - 1] \times 10.000$ (one basis point is equivalent to 0.01%). In this analysis 3-year moving average corresponds to our further minimum requirement for each fund, which is to have 36 months of continuous data to be included in the sample. This time period is commonly referred among academic literature (e.g. Cuthbertson et al., 2008; Petajisto 2013) An extension of the minimum requirement up to 60 months of continuous data would dramatically reduce the number of observations. Moreover, the results of Barras et al. (2010) show that reducing the minimum fund return requirement to 36 months has no material impact on the main results Authors state that any biases introduced from the 60-month requirement are minimal. The figure reveals persistent negative alpha of the index throughout the period. More extreme negative alpha coincides with the global financial crisis period of 2008-2010, while less extreme alpha values (and even a small positive FF3 alpha of 20bps) are recorded in the late 1990s, a period of dot.com boom and a strong bull market. These inconsistencies in non-zero alphas of FTSE 100 in different market states (bull vs. bear) lead to conjecture that mutual fund performance is more undervalued in bear markets and less undervalued (or overvalued in case of positive index alpha) in the bull markets.

Figure 1: FTSE100 alpha

The following regressions are estimated $R_{FTSE100,t,u} - R_{F,t,u} = \alpha_{FTSE100} + \beta_{M,t,u}(R_{M,t,u} - R_{F,t,u}) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + e_{t,u}$ and $R_{FTSE100,t,u} - R_{F,t,u} = \alpha_{FTSE100} + \beta_{M,t,u}(R_{M,t,u} - R_{F,t,u}) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + \beta_{WML}WML_{t,u} + e_{t,u}$ for the period for January 1992 to October 2013 (t is the frequency of the data, months and u represents the estimated subperiods in months). Monthly alpha is calculated for a three years (36 months) moving average (presented below in annual basis point). $R_{FTSE100,t,u} - R_{F,t,u}$ is the monthly excess return on the FTSE 100 index including dividends in period u, $R_{F,t,u}$ is the monthly risk-free rate in period u, α (alpha/constant) is the Fama-French and Carhart performance estimate, $(R_{M,t,u} - R_{F,t,u})$ is the monthly market risk premium in period u, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{t,u}$.is the monthly error term. The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website.

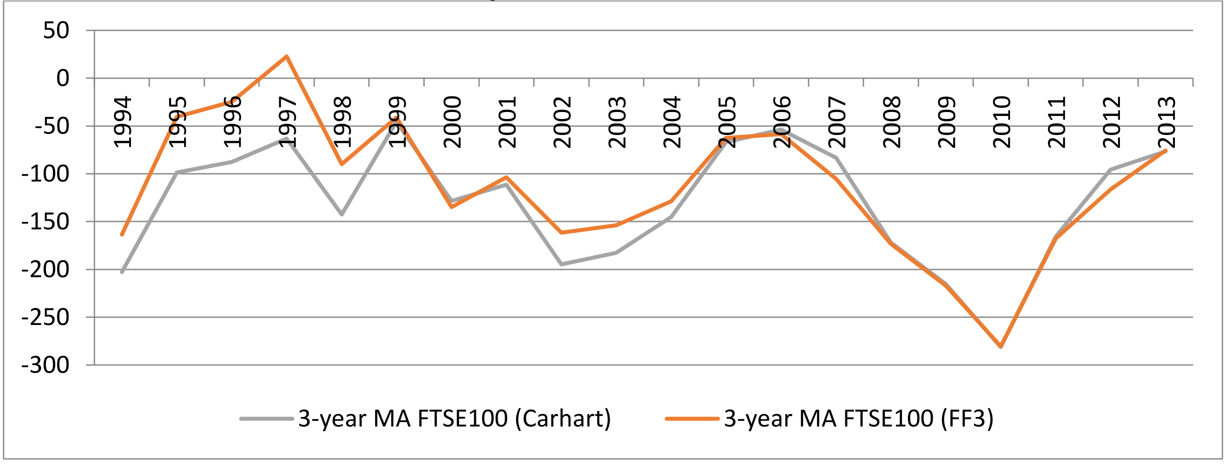


Table 3 takes a closer look at FTSE 100 performance in the overall sample period and in the bull and bear markets. Specifically, Table 3 lays out the FF3 and Carhart alphas of FTSE 100 index in the overall sample period and in five sub-periods. The FTSE 100 index generates a statistically significant *negative* FF3 alpha of -1.12% and Carhart alpha of -1.13% (both significant at 1% level) per annum for the entire sample period January 1992 – October 2013.

Moreover, non-zero annual alphas for the index have larger negative values in bear markets (ranging from -1.61% to -2.86%) then in bull markets (-0.47% and -1.10%). This difference is substantial, therefore being of economic significance to investors. Negative FTSE100 index alphas from Table 3 infer that the performance of UK funds benchmarking against FTSE 100 will be undervalued by the standard FF3 and Carhart models[6]. Such underperformance will particularly be amplified in bear markets. Once the models are modified to correct for the presence of negative benchmark index alpha, we expect the adjusted mutual fund alphas in bear markets to shift *upwards*.

Table 3: FTSE 100 Index Alpha regressions

The table reports alpha (intercept) per month and per year (in bps) from the following three- and four-factor model regressions: $FTSE_{TR_{t,u}} - RF_{t,u} = \alpha + \beta_{Mt,u}\left(RM_{t,u} - R_{Ft,u}\right) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + e_{t,u}$ and $FTSE_{TR_{t,u}} - RF_{t,u} = \alpha + \beta_{Mt,u}\left(RM_{t,u} - R_{Ft,u}\right) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + \beta_{WML}WML_{t,u} + e_{t,u}$ for the period for January 1992 to October 2013 (t is the frequency of the data, months and u represents the estimated subperiods in months). $FTSE_{TR_{t,u}} - RF_{t,u}$ is FTSE100 monthly total return in excess to the risk free rate. RF is the monthly risk free rate. α (alpha) is the constant term and $e_{t,u}$.the error term. P-values in parenthesis. Superscript *indicate statistical significance at 1% (***), 5% (**) and 10% (*) levels. Alphas are reported for the total sample period (January 1992-October 2013) and for five sub-periods. The number of months and adjusted R-squared from the three-factor and four-factor model are also reported.

Period and Model used	**Alpha per month**	**Alpha p.a. in bps**	**Market Beta**	**SMB**	**HML**	**WML**	**Number of Months**	**Adj. R-squared**
Total Sample	-0.0009383***	-112.017***	1.01***	-0.132***	-0.017***	---	262	0.9944
FF3	(0.000)		(0.000)	(0.000)	(0.002)			
Total Sample	-0.0009476***	-113.121***	1.01***	-0.132***	-0.017**	0.001	262	0.9944
Carhart	(0.000)		(0.000)	(0.000)	(0.013)	(0.871)		
1992:01 to 1999:12	-0.0003979	-47.6436	0.99***	-0.191***	-0.037***	---	96	0.9910
FF3	(0.301)		(0.000)	(0.000)	(0.000)			
1992:01 to 1999:12	-0.0005798	-69.3546	0.99***	-0.189***	-0.024	0.019	96	0.9911
Carhart	(0.157)		(0.000)	(0.000)	(0.129)	(0.204)		
2000:01 to 2002:12	-0.0013565**	-161.571**	1.02***	-0.107***	0.001	---	36	0.9965
FF3	(0.023)		(0.000)	(0.000)	(0.927)			
2000:01 to 2002:12	-0.0016378***	-194.775***	1.03***	-0.108***	0.015	0.019**	36	0.9969
Carhart	(0.005)		(0.000)	(0.000)	(0.203)	(0.029)		
2003:01 to 2007:12	-0.0008271***	-98.8017***	0.99***	-0.139***	-0.010	---	60	0.9967
FF3	(0.000)		(0.000)	(0.000)	(0.390)			
2003:01 to 2007:12	-0.0008051***	-96.1853***	0.99***	-0.139***	-0.011	-0.003	60	0.9966
Carhart	(0.001)		(0.000)	(0.000)	(0.360)	(0.712)		

[6] Note that the performance of funds benchmarking against indices whose alphas have positive values, the performance will be overstated.

2008:01 to 2009:12	-0.0024236***	-286.986***	1.03***	-0.087***	-0.050***	---	24	0.9989
FF3	(0.000)		(0.000)	(0.000)	(0.008)			
2008:01 to 2009:12	-0.0023247***	-275.425***	1.03***	-0.079***	-0.046**	0.009	24	0.9989
Carhart	(0.000)		(0.000)	(0.000)	(0.017)	(0.293)		
2010:01 to 2013:10	-0.0009215***	-110.021***	0.99***	-0.087***	0.006	---	46	0.9983
FF3	(0.001)		(0.000)	(0.000)	(0.646)			
2010:01 to 2013:10	-0.0007313***	-87.4039***	0.99***	-0.093***	-0.002	-0.014	46	0.9984
Carhart	(0.000)		(0.000)	(0.000)	(0.868)	(0.145)		

Prior UK studies show strong evidence of underperformance of UK mutual funds, as seen in Blake and Timmermann (1998), Quigley and Sinquefield (2000), Cuthbertson, Nitzsche and O'Sullivan (2008) among others. The negative FTSE 100 alphas from the standard FF3 and Carhart models in our study are at least in part covering the period of analysis in a number of these UK studies. Therefore, the use of misspecified performance evaluation models in these studies, which lead to negative benchmark alphas, may be the reason behind the evidence of persistent underperformance of UK mutual funds. It is then imperative that the UK mutual fund performance is re-assessed using the adjusted FF3 and Carhart models suggested in recent literature on performance measurement, such as Angelidis et. al. (2013).

3.2.3. Evaluating Mutual Fund Performance: Standard vs. Adjusted FF3 and Carhart Alphas

For each equity mutual fund *i* in our sample, we first estimate the standard FF3 factor and Carhart four-factor model alphas as per equations (3) and (4).

$$R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t} - R_{F,t}\right) + \beta_{iSMB}SMB_{t,u} + \beta_{i,HML}HML_{t,u} + e_{i,t,u} \quad (3)$$

$$R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t} - R_{F,t}\right) + \beta_{iSMB}SMB_{t,u} + \beta_{i,HML}HML_{t,u} + \beta_{i,WML}WML_{t,u} + e_{i,t,u} \quad (4)$$

where $R_{i,t,u}$ is the monthly return of a mutual fund *i* in period *u*, α_i is the monthly excess return of the fund *i* over period *u* and the rest of the variables are described as per equations (1) and (2).

Next, we apply Angelidis, Giamouridis and Tessaromatis (2013) adjustment to FF3 and Carhart model (AGT-adjustment hereafter) that accounts for benchmark index performance and consequently improves the accuracy of measuring the funds' excess returns:

$$R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1,t,u}^*\left(R_{M,t} - R_{F,t}\right) + \beta_{i2}^*SMB_{t,u} + \beta_{i3}^*HML_{t,u} + e_{i,t,u}^* \quad (5)$$

$$R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1,t,u}^*\left(R_{M,t} - R_{F,t}\right) + \beta_{i2}^* SMB_{t,u} + \beta_{i3}^* HML_{t,u} + \beta_{i4}^* WML_{t,u} + e_{i,t,u}^* \quad (6)$$

where $R_{i,t,u} - R_{FTSE100,t,u}$ is the monthly excess return of a mutual fund i over the FTSE 100 index in period u, α_i^* is the difference of the fund's and benchmark's FF3 (Carhart) alpha estimated in equations (3) and (1) (equations (4) and (2)); i.e. AGT-adjusted alpha hereafter. Additionally if the excess Beta ($\beta_{i1}^*, \beta_{i2,}^*\beta_{i3,}^*\beta_{i4}^*$) is different from zero (again obtained as the difference in betas between equation (3) and (1) or (4) and (2) in FF3 and Carhart model respectively) the manager has a portfolio in which beta differs from that of the FTSE 100. As an example, if the estimated SMB beta is 0.1 means that the fund's is 10 percent more exposed to small stocks than the benchmark.

3.3. Results

3.3.1. Standard FF3 and Carhart alpha of UK mutual funds

Using equations (3) and (4) and a fixed effects panel model estimation procedure we obtain standard FF3 and Carhart alphas for 887 funds in our sample[7]. The results of for five sub periods and the overall sample period are reported in Table 4.

Without the adjustment for the negative FTSE 100 alpha in the whole sample period, we find a positive annual FF3 alpha for our 887 equity mutual funds of 0.14% (13.81bps p.a., significant at 10% level) and a negative Carhart alpha of -0.29 % (28.76bps p.a., significant at 1%). In the sub-periods, funds exhibit higher standard alphas in the bull periods than in bear markets. The strongest positive alphas are recorded in the last bull period in our sample (2010-2013). However, while being statistically significant, they do not add great economic value to investors: FF3 alpha is 1.04% p.a. and Carhart only 0.69% p.a. in 2010-2013. With all WML coefficients throughout sub-periods being positive and significant at 1% level, there is evidence of strong managers' ability to successfully pick winner stocks and sell losers in their portfolios. In spite of this, funds' performance still results in a negative standard Carhart alpha in the overall period and most of the sub-periods. This further adds to the fact that if the benchmark index alphas are negative over the estimation period, the performance of funds benchmarking against that index is underestimated according to standard alpha models.

[7] Hausman test statistic was used to choose between the fixed and random effects estimation.

Hence, the performance estimates reported in Table 4 are not showing accurate reflection of UK equity mutual fund performance. Coefficient for SMB risk factor is positive in all sub-periods indicating presence of small cap risk in the funds, while the evidence on the presence of value/growth style risk is mixed (coefficients varying from positive to negative) across sub-periods. Section 4.3. will provide further insight into performance of funds in our sample by their investment style.

Table 4: Fixed Effects Panel FF3 and Carhart regressions for UK Equity Mutual Funds returns:

The sample consists in 887 unique UK Equity Mutual Funds and 123,768 monthly observations over the period January 1992 to October 2013 (t is the frequency of the data, months and u represents the estimated subperiods in months). The following regressions are estimated $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i,SMB}SMB_{t,u} + \beta_{i,HML}HML_{t,u} + e_{i,t,u}$ and $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i,SMB}SMB_{t,u} + \beta_{i,HML}HML_{t,u} + \beta_{i,WML}WML_{t,u} + e_{i,t,u}$. Monthly alpha is calculated for a five different time periods u: January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull market) and January 2000 to December 2002 and January 2008 to December 2009 (bear market). Alphas from benchmark index are collected from table 3. $R_{i,t,u} - R_{F,t,u}$ is the monthly excess return on equity mutual fund i for period u. $R_{F,t,u}$ is the monthly risk-free rate in period u, α (alpha/constant) is the Fama-French and Carhart performance estimate, $\left(R_{M,t,u} - R_{F,t,u}\right)$ is the monthly market risk premium in period u, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{i,t,u}$.is the monthly error term. The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website. P-values in parenthesis. Superscript *indicate statistical significanc
e at 1%(***), 5% (**) and 10% (*) levels.

Period	Alpha Measure	Equity Mutual Fund Alpha p.a. (in %)	Market Beta	SMB	HML	WML	Number Funds	Observations	Adj. R-squared
Total Sample	***FF3***	13.81* (0.057)	0.9357*** (0.000)	0.2355*** (0.000)	0.00066 (0.716)	---	887	123,768	0.7902
Total Sample	***Carhart***	-28.76*** (0.000)	0.94064*** (0.000)	0.24689*** (0.000)	0.02501*** (0.000)	0.03351*** (0.000)	887	123,768	0.7910
1992:01to 1999:12	***FF3***	32.45** (0.049)	0.9553*** (0.000)	0.27159*** (0.000)	-0.00955** (0.014)	---	385	26,107	0.7472
1992:01to 1999:12	***Carhart***	-17.99 (0.294)	0.95432*** (0.000)	0.28004*** (0.000)	0.02393*** (0.000)	0.04654*** (0.000)	385	26,107	0.7480
2000:01 to 2002:12	***FF3***	-83.68*** (0.0006)	0.89914*** (0.000)	0.25135*** (0.000)	0.02466*** (0.000)	---	504	15,690	0.7322
2000:01 to 2002:12	***Carhart***	10.81*** (0.000)	0.86898*** (0.000)	0.25316*** (0.000)	-0.01040* (0.060)	-0.04727*** (0.000)	504	15,690	0.7344
2003:01 to 2007:12	***FF3***	9.24 (0.3750)	0.91950*** (0.000)	0.25153*** (0.000)	-0.03234*** (0.000)	---	760	37,060	0.7499
2003:01 to 2007:12	***Carhart***	-51.48*** (0.000)	0.92068*** (0.000)	0.25957***	-0.00910*** (0.000)	0.06297*** (0.000)	760	37,060	0.7528

				(0.0540)					
2008:01 to 2009:12	***FF3***	-207.99***	0.97349***	0.22975***	-0.09078***	---	765	16,771	0.8374
		(0.000)	(0.000)	(0.000)	(0.000)				
2008:01 to 2009:12	***Carhart***	-92.01***	0.99426***	0.31484***	-0.04656***	0.09126***	765	16,771	0.8417
		(0.001)	(0.000)	(0.000)	(0.000)	(0.000)			
2010:01 to 2013:10	***FF3***	103.69***	0.93341***	0.23719***	0.07670***	---	735	28,140	0.8029
		(0.000)	(0.000)	(0.000)	(0.000)				
2010:01 to 2013:10	***Carhart***	68.62***	0.92961***	0.24637***	0.08997***	0.02150***	735	28,140	0.8031
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			

3.3.2. Adjusted FF3 and Carhart Alpha of UK mutual funds

This section reports AGT-adjusted FF3 and Carhart alphas for active UK equity mutual funds. The values of AGT-adjusted annualised alphas, the coefficients on the Market, SMB, HML and WML factor reported in Table 5 are obtained by estimating equations (5) and (6) with fixed effects panel model estimation, as in Section 4.1. For ease of comparison, in this table we also include the values of standard FF3 and Carhart annualised alphas previously reported in Table 4. Table 5 uniformly documents strong positive improvement in all FF3 (Panel A) and Carhart alphas (Panel B) after the AGT-adjustment in the whole sample period and each sub-period. The difference in standard and AGT-adjusted alphas is statistically significant at 1% level for the whole sample period and each of the sub-periods in both Panels; the exception is the first bull period 1992-1999, where the difference between standard and adjusted FF3 (Cahrart) alphas Panel A (Panel B) is significant at 10% (5%) level, as indicated by Z-test[8].

Specifically, in the total sample period, the value of annualised FF3 alpha increases more than tenfold from 14 to 144 bps (significant at 1% level[9]) when fund returns are benchmark-adjusted, using AGT model specification. Across sub-periods, the AGT-adjusted FF3 alphas are overall statistically significant and positive, which stands even in bear markets. The improvement in alphas post-adjustment ranges from 30bps in the first bull sub-period 1992-1999 to 289bp in the last bear period 2007-2009. What is more, FF3 alphas in bear markets change sign from negative (-84 bps in 2000-2002 and -208bps in 2008-2009) to positive (76bps and 81bps in the two bear periods respectively). Panel B shows qualitatively the same results for Carhart alpha adjustment. This is in line with our expectations that greater underestimation of fund performance in standard FF3 and Carhart models occurs in bear markets, due to presence of larger negative alphas of the benchmark index.

[8]Z-test is calculated as: $Z - stat = \frac{alpha_{before} - alpha_{adjusted}}{\sqrt{\left(SE_{alpha_{before}}\right)^2 + \left(SE_{alpha_{adjusted}}\right)^2}}$

[9] We have re-estimated standard errors of AGT-adjusted alphas in this paper using Petersen (2009) method and clustering by fund and months, fund and years and fund and bull/bear periods. Our alphas for the total sample of funds and funds per investment style by and large remain of the same level of significance.

Table 5: Fixed Effects Panel Data regressions for UK Equity Mutual Funds returns: FF3 and Carhart model alphas before and after AGT-adjustment with FTSE 100 benchmark

The sample consists of 887 unique UK Equity Mutual Funds and 123,768 monthly observations over the period January 1992 to October 2013(t is the frequency of the data, months and u represents the estimated subperiods in months). The following regressions are estimated: $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1,t,u}^*\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i2}^*SMB_{t,u} + \beta_{i3}^*HML_{t,u} + e_{i,t,u}^*$ (after adjustment) in Panel A; and $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + \beta_{WML,i}WML_{t.u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1,t,u}^*\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i2}^*SMB_{t,u} + \beta_{i3}^*HML_{t,u} + \beta_{i4}^*WML_{t,u} + e_{i,t,u}^*$ (after adjustment) in Panel B. $R_{i,t,u} - R_{F,t,u}$ is the monthly excess return on equity mutual fund *i* for period u. $R_{F,t,u}$ is the monthly risk-free rate in period u, α (alpha/constant) is the Fama-French and Carhart performance estimate, $\left(R_{M,t,u} - R_{F,t,u}\right)$ is the monthly market risk premium in period u, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{i,t,u}$.is the error term. α_i^* is the AGT-adjusted alpha and β_{i1}^*-β_{i4}^* are excess factor betas. This is done for the full time period (1992-2013), January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull market) and January 2000 to December 2002 and January 2008 to December 2009 (bear market). The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website. *P-values* are in parenthesis. Significance of the difference in alphas is determined by $Z - stat = \frac{alpha_{before} - alpha_{adjusted}}{\sqrt{\left(SE_{alpha_{before}}\right)^2 + \left(SE_{alpha_{adjusted}}\right)^2}}$. Superscript *indicate statistical significance at 1% (***), 5% (**) and 10% (*) levels.

Panel A: FF3 model and AGT-adjusted three factor model

	FF3 Alpha (annual basis points)				**Excess Market Beta**	**Excess SMB**	**Excess HML**	**Number of Funds**	**Obs.**	**R-Squared (within)**
	Before	***After AGT-adj.***	***Difference*** [10]	**Z-test**[11]						
Total Sample	13.81* (0.057)	143.64*** (0.000)	**129.83**	12.46***	-0.0741259*** (0.000)	0.3561225*** (0.000)	0.0129689*** (0.000)	887	123,768	0.7902/0.2368
1992:01-1999:12	32.45** (0.049)	62.54 (0.107)	**30.09**	1.68*	-0.0424562*** (0.000)	0.4623843*** (0.000)	0.0254355*** (0.000)	385	26,107	0.7472/0.2544
2000:01-2002:12	-83.68*** (0.0006)	76.54** (0.015)	**160.22**	3.65***	-0.1228128*** (0.000)	0.3571102*** (0.000)	0.0230574*** (0.000)	504	15,690	0.7322/ 0.2511
2003:01-2007:12	9.24 (0.3750)	112.06*** (0.000)	**102.82**	6.98***	-0.0734209*** (0.000)	0.3898085*** (0.000)	-0.0240685*** (0.000)	760	37,060	0.7499/ 0.3132
2008:01-2009:12	-207.99*** (0.000)	81.46*** (0.003)	**289.45**	7.62***	-0.0578245*** (0.000)	0.317387*** (0.000)	-0.0404989*** (0.000)	765	16,771	0.8374/0.2044

[10] The difference is calculated as alpha after the adjustment (column 2) minus alpha prior the adjustment (column 1).

[11] Note, that alphas used for Z-test are obtained from two different models Carhart and AGT, therefore the results provided in Z-test column should be treated with caution. Nevertheless, given the above shortcoming, it is clear that the AGT adjustments gives a parallel shift to the standard Carhart model and the results provided are of practical and economic relevance.

2010:01-2013:10	103.69***	217.40***	**113.71**	5.47***	-0.0555168	0.3241709***	0.0704108***	735	28,140	0.8029/0.1783
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)			

Panel B: Carhart model and AGT-adjusted four factor model

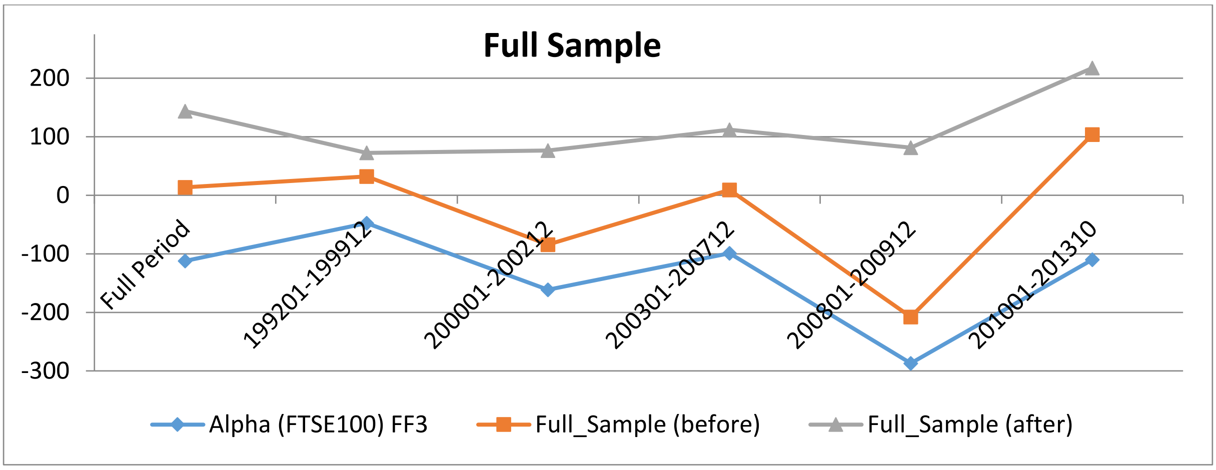
	Carhart Alpha (annual basis points)				**Excess Market Beta**	**Excess SMB**	**Excess HML**	**Excess WML**	**Number of Funds**	**Obs.**	**R-Squared (within)**
	Before	***After AGT-adj.***	***Difference***	**Z-test**							
Total Sample	-28.76***	98.57***	**127.33**	11.91***	-0.069***	0.368***	0.038***	0.035***	887	123,768	0.7910/ 0.2399
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)	(0.000)			
1992:01-1999:12	-17.99	35.58**	**53.57**	2.16**	-0.043***	0.469***	0.048***	0.034***	385	26,107	0.7480/0.2596
	(0.294)	(0.047)			(0.000)	(0.000)	(0.000)	(0.000)			
2000:01-2002:12	10.81***	210.10***	**199.29**	4.36***	-0.165***	0.356***	-0.025***	-0.065***	504	15,690	0.7344/0.2727
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)	(0.000)			
2003:01-2007:12	-51.48***	46.79***	**98.27**	6.47***	-0.072***	0.398***	0.001	0.068***	760	37,060	0.7528/0.3287
	(0.000)	(0.000)			(0.000)	(0.000)	(0.852)	(0.000)			
2008:01-2009:12	-92.01***	188.79***	**280.80**	7.26***	-0.039***	0.394***	-0.001	0.082***	765	16,771	0.8417/ 0.2300
	(0.001)	(0.000)			(0.000)	(0.000)	(0.936)	(0.000)			
2010:01-2013:10	68.62***	158.71***	**90.09**	3.87***	-0.062***	0.340***	0.093***	0.036***	735	28,140	0.8031/0.1847
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)	(0.000)			

Finally, similarly to Table 4, we observe positive and statistically significant coefficients for SMB in Table 5, indicating a small cap orientation of funds in the sample and the mixed results for HML risk factor coefficients. The performance of funds by styles will be addressed in the next section.

Figure 2 presents a summary of these results. Equity mutual fund alphas (before and after AGT-adjustment) are estimated for the full sample of 887 equity mutual funds for the full sample period and five sub-sample periods (bull and bear market). For comparative purpose the results are plotted versus benchmark alphas (FTSE 100), estimated in a similar matter, for the full-time period and five sub-periods. The Figure shows annualized FF3 alphas[12] in basis points, the following formula has been applied $[(1 + monthly\ alpha)^{12} - 1] \times 10.000$ (one basis point is equivalent to 0.01%).

Figure 2: Equity mutual fund (before and after AGT- adjustment) and FTSE 100 index alphas

Figure shows for different time periods the FTSE 100 index alpha (FF3 model), the equity mutual fund alpha before and after adjustment the non-zero benchmark index alpha. This is done for the full time period (1992-2013), January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull market) and January 2000 to December 2002 and January 2008 to December 2009 (bear market).



This illustration distinctly shows that after AGT model adjustment, fund alphas considerably improve. On the average, active UK equity mutual funds are able to generate positive outperformance without major declines even during the last financial crisis. Our findings corroborate our initial notion that UK equity mutual funds generate better performance than

[12] As the results for Carhart alphas are qualitatively the same, we do not report them in a separate figure

previously estimated in the literature deploying standard factor models for evaluating performance. Our results are also in line with Angelidis et al (2013) who report less negative and less statistically significant adjusted-alphas across their categories of funds, i.e. a better US mutual fund performance than the literature suggests.

3.3.3. Performance by Investment Styles

To test performance by funds style, we place each of the 887 mutual funds into one of the Morningstar style box categories: small-value, small-growth, small blend, large-value, large-growth, large-blend, mid cap-value, mid cap-growth and mid-cap blend. To identify the style category each fund should be placed to, we run individual regressions for each fund as specified by Equation (4). We then split the total sample of the funds according to their style characteristics given by SMB and HML coefficients from equation (3)[13]. There is 618 Small/Value, Small/Growth and Small/Blend style funds, representing 70 percent of the whole sample of 887 funds. Medium/Value and Medium/Growth comprise 6.7% of funds (59 out of 887), while there are 159 Medium/Blend funds, accounting for almost 18% of the total number. For each category of funds, we estimate FF3 and Carhart alphas before and after the AGT-adjustment using fixed effects panel estimation.

Table 6 lays out these results for the whole sample period. Panel A presents FF3 alphas and AGT-adjusted three factor model alphas, while Panel B shows Carhart alphas and AGT-adjusted four factor model alphas. Both panels report the differences in alphas, the significance of those differences (Z-test) and the Market, SMB, HML (and WML, in Panel B only) coefficients from the AGT- adjusted models.

[13] Small/value group: β SMB positive and statistically significant, β HML positive and statistically significant. Small/Growth group: β SMB positive and statistically significant, β HML negative and statistically significant. Small/blend group: β SMB positive and statistically significant, β HML not statistically significant. Large/value group: β SMB negative and statistically significant, β HML positive and statistically significant. Large/growth group: β SMB negative and statistically significant, β HML negative and statistically significant. Large/blend: β SMB negative and statistically significant, β HML not statistically significant. Medium/value group: β SMB not statistically significant, β HML positive and statistically significant. Medium /growth group: β SMB not statistically significant, β HML negative and statistically significant. Medium/blend group: β SMB not statistically significant, β HML not statistically significant.

Table 6: UK Equity Mutual Funds alphas by fund style, before and after AGT-adjustment: total sample period Jan 1992 – Oct 2013

887 equity mutual funds are divided styles as per Morningstar style box (Small/Value, Small/Growth, Small/Blend, Medium/Value, Medium/Growth, Medium/Blend, Large/Value, Large/Growth, Large/Blend). The following regressions are estimated: $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1,t,u}^*\left(R_{M,t} - R_{F,t}\right) + \beta_{i2}^*SMB_{t,u} + \beta_{i3}^*HML_{t,u} + e_{i,t,u}^*$ (after adjustment) in Panel A; and $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + \beta_{WML,i}WML_{t,u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE100,t,u} = \alpha_i^* + \beta_{i1.t,u}^*\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i2}^*SMB_{t,u} + \beta_{i3}^*HML_{t,u} + \beta_{i4}^*WML_{t,u} + e_{i,t,u}^*$ (after adjustment) in Panel B. $R_{i,t,u} - R_{F,t,u}$ is the monthly excess return on equity mutual fund *i* for period *u* (*t* is the frequency of the data, months and *u* represents the estimated subperiods in months). $R_{F,t,u}$ is the monthly risk-free rate in period *u*, α (alpha/constant) is the Fama-French and Carhart performance estimate, $\left(R_{M,t,u} - R_{F,t,u}\right)$ is the monthly market risk premium in period *u*, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{i,t,u}$.is the error term. α_i^* is the AGT-adjusted alpha and β_{i1}^*-β_{i4}^* are excess factor betas. The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website. *P-values* are in parenthesis. Significance of the difference in alphas is determined by $z - stat = \frac{alpha_{before} - alpha_{adjusted}}{\sqrt{\left(SE_{alpha_{before}}\right)^2 + \left(SE_{alpha_{adjusted}}\right)^2}}$. Superscript *indicate statistical significance at 1% (***), 5% (**) and 10% (*) levels.

Panel A: FF3 model and AGT-adjusted three factor model

Investment Style	**FF3 Alpha (annual basis points)**				**Excess Market Beta**	**Excess SMB**	**Excess HML**	**Number of Funds**	**Obs.**	**R-Squared (within)**
	Before	*After AGT-adj.*	*Difference*[14]	**Z-test**						
Small_Value	40.42***	162.51***	**122.09**	5.86***	-0.093***	0.396***	0.153***	135	26,764	0.8186/0.3212
					(0.000)	(0.000)	(0.000)			
Small_Growth	81.56***	203.6***	**122.04**	4.20***	-0.077***	0.581***	-0.171***	118	22,004	0.7577/0.4161
					(0.000)	(0.000)	(0.000)			
Small_Blend	14.27	153.96***	**139.69**	8.26***	-0.067***	0.391***	0.016***	365	44,031	0.8110/0.2841
					(0.000)	(0.000)	(0.000)			
Large_Value	-148.5***	-15.12	**133.38**	2.74***	-0.017***	0.038***	0.056***	6	925	0.9597/ 0.0679
					(0.007)	(0.000)	(0.000)			
Large_Growth	-26.15	98.31	**124.46**	0.83	-0.061***	0.011	-0.113***	5	659	0.7515/ 0.0461
					(0.005)	(0.671)	(0.000)			
Large_Blend	-94.72***	29.14	**123.86**	3.28***	-0.048***	0.029***	-0.006	40	5,779	0.8515/0.0166
					(0.000)	(0.000)	(0.314)			
Medium_Value	24.80	144.24***	**119.44**	3.64***	-0.095	0.144***	0.127***	35	6,324	0.8727/0.1607
					(0.000)	(0.000)	(0.000)			
Medium_Growth	2.62	123.17***	**120.55**	2.50***	-0.054***	0.141***	-0.084***	24	2,879	0.8733/ 0.1390
					(0.000)	(0.000)	(0.000)			
Medium_Blend	-83.34***	-48.59***	**34.75**	5.81***	-0.068***	0.133***	0.003	159	14,403	0.8608/0.0856
					(0.000)	(0.000)	(0.462)			

[14] The difference is calculated as alpha after the adjustment (column 2) minus alpha prior the adjustment (column 1).

Panel B: Carhart model and AGT-adjusted four factor model

Investment Style	**Carhart Alpha (annual basis points)**				**Excess Market Beta**	**Excess SMB**	**Excess HML**	**Excess WML**	**Number of Funds**	**Obs.**	**R-Squared (within)**
	Before	*After AGT-adj.*	*Difference*	**Z-test**							
Small_Value	43.19***	165.34***	**122.15**	5.65***	-0.093*** (0.000)	0.395*** (0.000)	0.152*** (0.000)	-0.002 (0.495)	135	26,764	0.8186/0.3212
Small_Growth	-22.50	98.71***	**121.21**	4.10***	-0.063*** (0.000)	0.603*** (0.000)	-0.115*** (0.000)	0.077*** (0.000)	118	22,004	0.7616/0.4249
Small_Blend	-39.04***	96.65***	**135.69**	7.85***	-0.061*** (0.000)	0.410*** (0.000)	0.049*** (0.000)	0.047*** (0.000)	365	44,031	0.8123/0.2898
Large_Value	-120.85***	14.09	**134.94**	2.69***	-0.029*** (0.001)	0.032*** (0.000)	0.041*** (0.000)	-0.021 (0.002)	6	925	0.9601/0.0780
Large_Growth	-71.97	52.43	**124.40**	0.81	-0.056*** (0.010)	0.021 (0.455)	-0.089*** (0.005)	0.034 (0.143)	5	659	0.7524/0.0492
Large_Blend	-101.64***	23.12	**124.76**	3.19***	-0.047*** (0.000)	0.031*** (0.000)	-0.003 (0.692)	0.004 (0.407)	40	5,779	0.8515/0.0167
Medium_Value	22.76	143.20***	**120.44**	3.53***	-0.095*** (0.000)	0.145*** (0.000)	0.127*** (0.000)	0.001 (0.878)	35	6,324	0.8727/0.1607
Medium_Growth	-51.78	67.31*	**119.09**	2.41***	-0.0476581 (0.000)	0.154*** (0.000)	-0.053 (0.000)	0.042 (0.000)	24	2,879	0.8748/0.1492
Medium_Blend	-109.58***	19.05	**128.63**	5.49***	-0.065*** (0.000)	0.142*** (0.000)	0.020*** (0.000)	0.0237 (0.000)	159	14,403	0.8611/0.0885

Table 6 demonstrates that results per fund category are consistent with the overall sample of funds from Table 3. In both Panel A and Panel B, the AGT-adjustment leads to improvement in alphas in each style category over the sample period. The differences in standard and AGT-adjusted alphas are significant at 1% level for all fund categories except Large/Growth. The AGT-adjusted alphas are positive and statistically significant for all Small Cap sub-categories as well as Medium/Value and Medium/Growth groups. According to standard FF3 (Panel A) and Carhart (Panel B) models, large cap funds generate negative alphas and underperform other fund styles. After the AGT-adjustment, large cap funds performance is in line with the market. In Panel A, the best performing group are small cap growth funds with 82bps in the standard FF3 model and 204bps AGT-adjusted three factor alpha per year. According to Carhart alphas in Panel B, small cap value funds are best performing with 43bps standard and 165bp AGT-adjusted alpha. In general, in both panels, small cap funds outperform the medium or large cap funds in each of the corresponding sub-categories ('value', 'growth' and 'blend'). They generate positive ATG-adjusted alphas across all subcategories, significant at 1% level. It is interesting to note that 'blend' funds generate overall negative performance according to standard performance measures, which is consistent with Jennifer, Sialm, and Zhang (2011) who provide evidence that funds that tend to shift risks perform worse than others. Once AGT-adjustment is applied, small/blend and medium/blend categories in Panel B exhibit greatest increase in alphas within their size categories, which turn from negative to positive values. In spite of this strong improvement in adjusted alphas, 'blend' funds do not perform as well as the value and growth group within the same size category.

We now turn our analysis to the bull and bear market sub-periods, which will help us identify if the performance of some style groups is driven by any particular sub-period. Table 7 reports FF3 and Carhart alphas before and after the AGT-adjustment in the five sub-periods. The table is separated into four panels to differentiate between three and four factor models on the one hand, and bull and bear periods[15] on the other. Results for the FF3 model are presented in Panel A for the bull and Panel B for the bear periods; while results for the Carhart four factor model are in Panel C for the bull and Panel D for the bear periods.

[15] Bull and bear periods are defined in section 2 of this paper

Table 7: Annualized (in bps) bull vs. bear market FF3 and Carhart alphas before and after AGT-adjustment, per investment style category

Panel A reports standard FF3 and AGT-adjusted FF3 alphas and their difference in bull periods January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013; Panel B reports the same for the bear market periods January 2000 to December 2002 and January 2008 to December 2009. Panel C reports standard Carhart, AGT-adjusted Carhart alphas and their difference in the bull market periods and Panel D reports the same for the bear market. All alphas and their difference are annualized values in basis points. Results in each panel are presented by "investment style" category (Small/Value, Small/Growth, Small/Blend, Medium/Value, Medium/Growth, Medium/Blend, Large/Value, Large/Growth, Large/Blend) following Morningstar Equity Style box allocation. Significance of the difference in alphas is determined by $Z-stat = \frac{alpha_{before}-alpha_{adjusted}}{\sqrt{\left(SE_{alpha_{before}}\right)^2+\left(SE_{alpha_{adjusted}}\right)^2}}$. ***indicates statistical significance at 1%, ** at 5% and * at 10% level.

Panel A: FF3 and AGT-adjusted three factor model alpha in <u>bull</u> market periods

Investment Style	*1992:01-1999:12*				**Funds /Obs.**	*2003:01-2007:12*				**Funds /Obs.**	*2010:01-2013:10*				**Funds /Obs.**
	Before	**After AGT-adj.**	**Difference** [16]	**Z-test**		**Before**	**After AGT-adj.**	**Difference**	**Z-test**		**Before**	**After AGT adj.**	**Difference**	**Z-test**	
Small_Value	-18.90	25.00	**43.9**	1.18	104/7,758	25.42	126.40***	**100.98**	3.08***	126/7,121	131.11***	245.90***	**114.79**	2.35***	122/5,044
Small_Growth	174.03***	212.65***	**38.62**	0.62	85/5,893	36.87	136.26***	**99.39**	2.51***	115/6,460	216.00***	332.06***	**116.06**	1.82*	97/3,860
Small_Blend	43.66	84.38*	**40.72**	0.90	93/5,930	38.03**	143.80***	**105.77**	4.27***	321/14,274	119.64***	233.41***	**113.77**	3.48***	311/12,213
Large_Value	-88.50	-68.75	**19.75**	0.13	5/147	-216.69***	-120.36***	**96.33**	1.74*	6/301	-221.10*	-104.79	**116.31**	0.68	5/177
Large_Growth	105.26	147.63	**42.37**	0.17	2/175	-130.67	-26.21	**104.46**	0.87	4/203	-75.95***	34.34	**110.29**	3.34***	3/121
Large_Blend	-100.78*	-80.10	**20.68**	0.27	24/1,138	-149.35***	-48.26*	**101.09**	2.86***	37/1,759	6.91	106.94***	**100.03**	1.87*	31/1,176
Medium_Value	-76.70**	-6.19	**70.51**	0.80	26/1,975	-23.53	77.60**	**101.13**	2.06**	33/1,693	7.53	123.40**	**115.87**	1.34	28/1,038
Medium_Growth	-43.60	0.00	**43.6**	0.48	10/770	-6.55	92.24*	**98.79**	1.25	15/765	111.21	223.32***	**112.11**	1.00	20/656
Medium_Blend	-76.70**	-42.29	**34.41**	0.61	36/2,321	-64.54***	38.66*	**103.2**	3.40***	103/4,484	9.56	119.34***	**109.78**	2.35***	118/3,855

[16] The difference is calculated as alpha after the adjustment (column 2) minus alpha prior the adjustment (column 1).

Panel B: FF3 and AGT-adjusted three factor model alpha in bear market periods

Investment Style	*2000:01-2002:12*				**Funds /Obs.**	*2008:01-2009:12*				**Funds /Obs.**
	Before	**After AGT-adj.**	**Difference**	**Z-test**		**Before**	**After AGT-adj.**	**Difference**	**Z-test**	
Small_Value	-44.31	118.33**	**162.64**	2.14***	112/3,862	-58.44	235.17***	**293.61**	3.35***	132/2,979
Small_Growth	83.39	247.00***	**163.61**	1.48	99/3,323	-243.48***	44.24	**287.72**	2.74***	107/2,468
Small_Blend	-147.09***	10.96	**158.05**	2.08**	164/4,388	-231.53***	57.19	**288.72**	5.17***	326/7,226
Large_Value	-82.53	80.23	**162.76**	1.68*	5/180	34.29	329.96***	**295.67**	2.13**	5/120
Large_Growth	-237.19	-77.35	**159.84**	0.67	3/86	-256.79	34.54	**291.33**	0.24	4/74
Large_Blend	-327.53***	-173.69***	**153.84**	1.79*	27/934	-109.96	181.72	**291.68**	1.30	35/772
Medium_Value	43.59	208.98**	**165.39**	1.32	28/969	-237.30***	-11.93	**225.37**	2.18**	29/649
Medium_Growth	-104.98	57.44	**162.42**	0.96	10/360	-365.44***	-78.23	**287.21**	1.52	16/328
Medium_Blend	-249.31***	-91.44	**157.87**	1.85*	56/1,588	-337.818***	-52.09	**285.728**	3.32***	111/2,155

Panel C: Carhart and AGT-adjusted four factor model alpha in bull market periods

Investment Style	*1992:01-1999:12*				**Funds /Obs.**	*2003:01-2007:12*				**Funds /Obs.**	*2010:01-2013:10*				**Funds /Obs.**
	Before	**After AGT-adj.**	**Difference**	**Z-test**		**Before**	**After AGT-adj.**	**Difference**	**Z-test**		**Before**	**After AGT-adj.**	**Difference**	**Z-test**	
Small_Value	-14.70	45.82	**60.52**	1.53	104/7,758	5.46	103.13***	**97.67**	2.88***	126/7,121	135.61***	226.54***	**90.93**	1.68*	122/5,044
Small_Growth	-8.19	43.41	**51.6**	0.81	85/5,893	-60.72**	34.55	**95.27**	2.47***	115/6,460	98.80**	188.98***	**90.18**	1.29	97/3,860
Small_Blend	-3.05	49.75	**52.8**	1.11	93/5,930	-37.41**	61.64***	**99.05**	3.89***	321/14,274	70.67***	160.29***	**89.62**	2.48***	311/12,213
Large_Value	-107.42	-81.02	**26.40**	0.17	5/147	-191.41***	-97.46**	**93.95**	1.65*	6/301	-230.36*	-137.69	**92.67**	0.49	5/177
Large_Growth	137.09	194.88	**57.79**	0.22	2/175	-187.90**	-87.77	**100.13**	0.81	4/203	-72.82***	19.84	**92.66**	2.56***	3/121
Large_Blend	-93.00*	-66.06	**26.94**	0.34	24/1,138	-175.36***	-77.75***	**97.61**	2.68***	37/1,759	37.03	126.65***	**89.62**	1.32	31/1,176
Medium_Value	-4.12	56.04	**60.16**	1.06	26/1,975	-38.54	59.32*	**97.86**	1.92**	33/1,693	6.99	99.38	**92.39**	0.96	28/1,038
Medium_Growth	-72.63	-9.48	**63.15**	0.65	10/770	-89.92	4.12	**94.04**	1.18	15/765	129.01	218.60***	**89.50**	0.71	20/656
Medium_Blend	-68.90*	-22.97	**45.93**	0.78	36/2,321	-124.47***	-27.21	**97.26**	3.13***	103/4,484	7.68	93.45***	**85.77**	1.66*	118/3,855

Panel D: Carhart and AGT-adjusted four factor model alpha in bear market periods										
Investment Style	***2000:01-2002:12***				**Funds /Obs.**	***2008:01-2009:12***				**Funds /Obs.**
	Before	**After AGT-adj.**	**Difference**	**Z-test**		**Before**	**After AGT-adj.**	**Difference**	**Z-test**	
Small_Value	42.53	241.53***	**199.00**	2.55***	112/3,862	-18.19	264.32***	**282.51**	3.14***	132/2,979
Small_Growth	205.39***	407.34***	**201.95**	1.76*	99/3,323	-42.18	238.76***	**280.94**	2.67***	107/2,468
Small_Blend	-28.03	171.72***	**143.69**	2.51***	164/4,388	-92.38**	188.18***	**280.56**	4.95***	326/7,226
Large_Value	-45.34	152.16**	**106.82**	2.04**	5/180	-11.44	270.82***	**282.26**	2.06***	5/120
Large_Growth	-189.79	11.59	**201.38**	0.80	3/86	-202.93	79.45	**282.38**	0.23	4/74
Large_Blend	-268.64***	-78.57	**190.07**	2.17**	27/934	-71.33	208.99	**280.32**	1.22	35/772
Medium_Value	106.77	307.70***	**200.93**	1.56	28/969	-276.40***	-2.43	**273.97**	2.05**	29/649
Medium_Growth	-84.25	112.54	**196.79**	1.13	10/360	-198.75	80.94	**279.69**	1.50	16/328
Medium_Blend	-177.83***	19.86	**197.69**	2.22**	56/1,588	-230.39***	46.60	**276.99**	3.18***	111/2,155

The differences in the standard and the AGT-adjusted alphas in Panels A-D show that the standard three factor model undervalues fund performance in bear periods more than in bull periods, corroborating our findings from Table 3. This is particularly pronounced during the last bear period in the sample corresponding to the most recent financial crisis, January 2008 – December 2009, where the standard FF3 and Carhart models underestimate performance compared to the AGT-adjusted model by well over 2% per year in all fund categories. There is greater difference in standard and the AGT-adjusted alphas in the later rather than earlier sub-periods in the sample. Also, there is greater significance in the difference in alphas documented by Z-test in the later periods in our sample. The difference in alphas is most persistently significant for all small cap fund groups, large/value and medium/blend categories across all four panels in Table 7. In the first bull period 1992-1999, the differences in alphas are the smallest and not statistically significant, which is reflecting our findings for the whole sample of funds from Table 5.

Looking at the small size category performance over sub-periods, it can be said that most consistent outperformance across sub-periods according to adjusted alphas is in the small cap/value group, in line with numerous empirical evidence documenting outperformance of small cap and value stocks[17]. The highest AGT-adjusted FF3 alpha over the whole sample period (Panel A, Table 6), generated by the Small/growth category is largely driven by a dot.com boom in the 1990s and the most recent post-crisis period (Panels A and B, Table 7). Small/blend funds have competitive advantage in bull markets according to Table 7. Within the large size category that in the overall period does not generate significant alphas, we note that Large cap/Value group generates particularly large annual AGT-adjusted FF3 alpha of 3.29% (Carhart equivalent of 2.71%) during the latest financial crisis. This implies that investors' tendency for 'flight to safety' in turbulent periods, i.e. investment in larger companies that pay dividends, is justified. Medium/Value category generates the highest positive alphas in the aftermath of the dot.com boom (January 2000-December 2003), while Medium/Growth and Medium/Blend categories do best in the aftermath of the recent financial crisis (January 2010 – October 2013).

In conclusion to this section, our most significant finding arises from the fact that assessing UK equity mutual fund performance using adjusted FF3 and Carhart model that corrects for

[17] For the UK evidence, see for instance Dimpson and Marsh (2001), Levis (1985), Levis and Liodakis (1999).

the ‘errors’ in alphas in the original versions of those models leads us to conclude that UK equity funds have actually performed better than suggested by the existing literature.

3.4. Robustness check: FTSE Small Cap Index as a benchmark

In this paper we have used FTSE 100 Index as a benchmark for all the funds in our sample. One may argue that using fund-style-specific benchmarks will be more appropriate, but unfortunately, benchmarks accounting for combinations of styles such as small/growth, medium/value index etc. are not available. Therefore, in this section we replicate the methodology and present findings for the subset of UK equity mutual funds that were categorised as Small Cap (including all three sub-categories: Value, Growth and Blend) in the analysis in section 4.3. A total of 618 funds was identified, representing 69.7% of our total sample. We benchmark the performance of those funds against a more appropriate index given their style category – the FTSE Small Cap Index. Total returns of FTSE Small Cap index (inclusive of dividends) are from Datastream. We note that FTSE 100 and FTSE Small Cap Index are highly related, having a correlation coefficient of 0.76 over our sample period. This section reports results equivalent to Figure 1, Figure 2 and Table 5 from section 4.1 and 4.2[18].

Figure 3 illustrates three-year moving average of FF3 and Carhart alphas for FTSE Small Cap index. The benchmark alphas are obtained for 22 sub-periods starting from January 1992 to December 1994 as a first time period (displayed as the year 1994 on the Figure 3), January 1993-December 1995 as a second and thus continues until the final slot from January 2011 to October 2013. The alpha values (estimated with FF3 and the Carhart model) are annualised and given in basis points following the formula $[(1 + monthly\ alpha)^{12} - 1] \times 10.000$ (one basis point is equivalent to 0.01%). As has been mentioned previously, in this analysis 3-year moving average corresponds to our minimum requirement for each fund, which is to have 36 months of continuous data to be included in the sample. This time period is commonly referred among academic literature (e.g. Cuthbertson et al., 2008; Petajisto 2013) An extension of the minimum requirement up to 60 months of continuous data would dramatically reduce the number of observations. Moreover, the results of Barras et al. (2010) show that reducing the minimum fund return requirement to 36 months has no material impact on the main results Authors state that any biases introduced from the 60-month requirement are minimal.

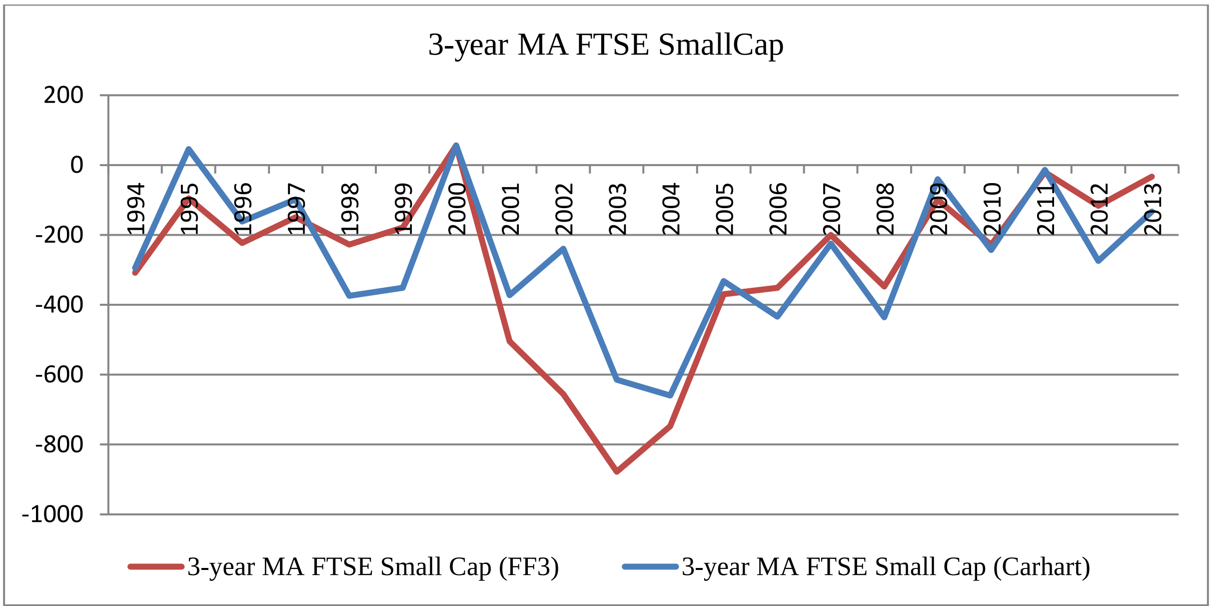
[18] Note that equivalents of the remaining tables, i.e. Tables 2, 4 and 5 are available on request from authors but are not reported here due to space constraints.

The results show that both alphas indicate even more pronounced underperformance of FTSE Small Cap Index relative to that of FTSE 100, reported in Figure 1. The sharpest decrease in alpha corresponds to dot-com bubble burst period in our sample, 2000-2003 (lowest recoded value is -878bps); while the only period of small positive alphas (56bps) was the dot-com boom period. This implies that in the same manner as with FTSE 100 as a benchmark, adjusting fund performance for index underperformance is expected to produce an upward shift in ATG-adjusted alphas for the small cap funds.

Table 8, in which FTSE Small Cap Index is used for AGT adjustment, corroborates those expectations. Panel A of Table 8 shows the results of the fixed effects panel model used to obtain standard FF3 and AGT-adjusted alphas given in basis points per annum, their difference, significance of the difference; the market, SMB and HML AGT-coefficients, number of funds, number of observations and model's R-squared. Panel B reports the Carhart model equivalents. The table is corresponding to Table 5, where FTSE 100 was used as a benchmark. The results in Table 8 are consistent and even more convincing than those reported in Table 5.

Figure 3: FTSE Small Cap alpha

The following regressions are estimated $R_{FTSE\ Small,t,u} - R_{F,t,u} = \alpha_{FTSE\ Small} + \beta_{M,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + e_{t,u}$ and $R_{FTSE\ Small,t,u} - R_{F,t,u} = \alpha_{FTSE\ Small,t,u} + \beta_{M,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB}SMB_{t,u} + \beta_{HML}HML_{t,u} + \beta_{WML}WML_{t,u} + e_{t,u}$ for the period for January 1992 to October 2013 (t is the frequency of the data, months and u represents the estimated subperiods in months). Monthly alpha is calculated for a three years (36 months) moving average (presented below in annual basis point). $R_{FTSE\ Small,t,u} - R_{F,t,u}$ is themonthly excess return on the FTSE Small Cap index including dividends in period u, $R_{F,t,u}$ is the monthly risk-free rate in period u, α (alpha/constant) is the Fama-French and Carhart performance estimate, $\left(R_{M,t,u} - R_{F,t,u}\right)$ is the monthly market risk premium in period u, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{t,u}$.is the error term. The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website



AGT-adjustment generates alphas significantly above standard FF3 and Carhart model estimates. This is consistent both in the overall sample period and all sub-periods. FF3 alpha increase post adjustment ranges from 53bps (period 2010-2013) to 759 bps (period 2000-2002). AGT-adjusted Carhart alpha shows improvement in performance between 81bps (period 2008-2009) and 441bps (1992-1999). In the total sample period, the FF3 alpha increases tenfold by 339 bps, while Carhart alpha exhibits rise of 291bps post AGT-adjustment. Z-tests shows that the differences in alphas are statistically significant across both Panels of Table 8, mostly at 1% level. This re-iterates that our results from section 4 are robust to the choice of the benchmark index.

Table 8: Fixed Effects Panel Data regressions for UK Equity Mutual Funds returns: FF3 and Carhart model alphas before and after AGT-adjustment with FTSE Small Cap benchmark

The sample consists in 618 unique UK Equity Mutual Funds and 92,799 monthly observations over the period January 1992 to October 2013 (t is the frequency of the data, months and u represents the estimated subperiods in months). The following regressions are estimated: $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE\ Small,t,u} = \alpha_i^{*} + \beta_{i1,t,u}^{*}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i2}^{*}SMB_{t,u} + \beta_{i3}^{*}HML_{t,u} + e_{i,t,u}^{*}$ (after adjustment) in Panel A; and $R_{i,t,u} - R_{F,t,u} = \alpha_i + \beta_{Mi,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{SMB,i}SMB_{t,u} + \beta_{HML,i}HML_{t,u} + \beta_{WML,i}WML_{t,u} + e_{i,t,u}$ (before adjustment) and $R_{i,t,u} - R_{FTSE\ Small,t,u} = \alpha_i^{*} + \beta_{i1}^{*}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i2}^{*}SMB_{t,u} + \beta_{i3}^{*}HML_{t,u} + \beta_{i4}^{*}WML_{t,u} + e_{i,t,u}^{*}$ (after adjustment) in Panel B. $R_{i,t,u} - R_{F,t,u}$ is the monthly excess return on equity mutual fund i for period u. $R_{FTSE\ Small,t,u}$ is the total return of FTSE Small Cap index in period u. $R_{F,t,u}$ is the monthly risk-free rate in period u, α (alpha/constant) is the Fama-French and Carhart performance estimate, $\left(R_{M,t,u} - R_{F,t,u}\right)$ is the monthly market risk premium in period u, SMB and HML are Fama and French (1993) size and value factors respectively, WML is Carhart (1997) momentum factor and $e_{i,t,u}$.is the error term. . α_i^{*} is the AGT-adjusted monthly alpha and β_{i1}^{*}-β_{i4}^{*} are excess factor betas. This is done for the full time period (1992-2013), January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull market) and January 2000 to December 2002 and January 2008 to December 2009 (bear market). The monthly risk factors and risk free rate are collected from University of Exeter, Xfi Centre for Finance and Investment website. *P-values* are in parenthesis. Significance of the difference in alphas is determined by $Z - stat = \frac{alpha_{before} - alpha_{adjusted}}{\sqrt{\left(SE_{alpha_{before}}\right)^2 + \left(SE_{alpha_{adjusted}}\right)^2}}$. Superscript *indicate statistical significance at 1% (***), 5% (**) and 10% (*) levels.

Panel A: FF3 model and AGT-adjusted three factor model

	FF3 Alpha (annual basis points)				Excess Market Beta	Excess SMB	Excess HML	Number of Funds	Obs.	R-Squared (within)
	Before	*After AGT-adj.*	*Difference*	Z-test						
Total Sample	37.14***	376.26***	**339.12**	25.88***	-0.0276403***	-0.4933105***	-0.0963244***	618	92,799	0.7859/ 0.3354
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)			
1992:01-1999:12	60.99***	353.42***	**292.43**	10.03***	0.038813***	-0.5184793***	-0.1438147***	282	19,581	0.7327/ 0.2964

	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)			
2000:01-2002:12	-48.16	710.53***	**758.70**	12.89***	-0.1357732***	-0.5265839***	-.01140015***	375	11,573	0.7219/ 0.3205
	(0.204)	(0.000)			(0.000)	(0.000)	(0.000)			
2003:01-2007:12	38.13***	272.47***	**234.34**	12.34***	-0.0169282***	-0.4536958***	-0.1175154***	562	27,855	0.7460/ 0.3475
	(0.002)	(0.000)			(0.000)	(0.000)	(0.000)			
2008:01-2009:12	-192.22***	-1.30	**190.92**	4.40***	-0.0775052***	-0.4550481***	-0.1870532***	565	12,673	0.8478/ 0.5161
	(0.000)	(-0.84)			(0.000)	(0.000)	(0.000)			
2010:01-2013:10	139.90***	193.41	**53.51**	2.05**	0.105324***	-0.4370582***	0.0417551***	530	21,117	0.7914/ 0.2533
	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)			

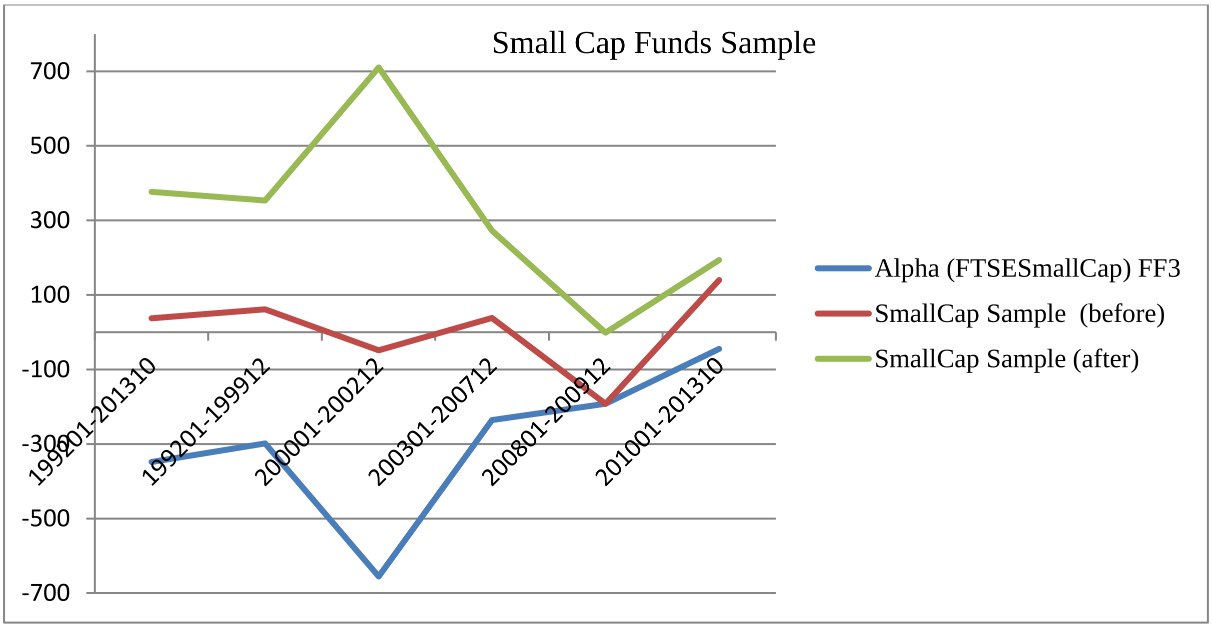
Panel B: Carhart model and AGT-adjusted four factor model

	Carhart Alpha (annual basis points)				Excess Market Beta	Excess SMB	Excess HML	WML differential	Number of Funds	Obs.	R-Squared (within)
	Before	*After AGT-adj.*	*Difference*	**Z-test**							
Total Sample	-14.62* (0.098)	275.54*** (0.000)	**290.16**	21.69***	-0.0164589*** (0.000)	-0.4669055*** (0.000)	-0.0404432*** (0.000)	0.07702*** (0.000)	618	92,799	0.7871/ 0.3463
1992:01-1999:12	-8.31 (0.692)	432.56*** (0.000)	**440.87**	14.34***	0.0402107*** (0.000)	-0.5308038*** (0.000)	-0.1940534*** (0.000)	-0.0700122*** (0.000)	282	19,581	0.7340/ 0.3008
2000:01-2002:12	59.80 (0.127)	320.32 (0.000)	**260.52**	4.46***	-0.0183653*** (0.000)	-0.5333997*** (0.000)	0.0228298*** (0.002)	0.1842752*** (0.000)	375	11,573	0.7245/ 0.3843
2003:01-2007:12	-28.14** (0.030)	201.47*** (0.000)	**229.61**	11.76***	-0.0155747*** (0.000)	-0.4443821*** (0.000)	-0.0907341 (0.000)	0.0724778*** (0.000)	562	27,855	0.7491/ 0.3546
2008:01-2009:12	-63.19** (0.040)	17.62 (0.581)	**80.81**	1.83*	-0.0741597*** (0.000)	-0.4414543*** (0.000)	-0.1800103*** (0.000)	0.0145683*** (0.004)	565	12,673	0.8528/ 0.5164
2010:01-2013:10	91.65*** (0.000)	275.31*** (0.000)	**183.66**	6.31***	0.1143765*** (0.000)	-0.4587158*** (0.000)	0.0102935 (0.257)	-0.0507207*** (0.000)	530	21,117	0.7917/ 0.2559

Figure 4 provides summary of the results for the sub-sample of 618 Small cap equity funds and is equivalent of Figure 2 from Section 4.2. Figure illustrates annualized FF3 alphas[19] (in bps) for the FTSE Small Cap Index and for the small-cap sub-sample of funds as well as annualized AGT-adjusted alphas (in bps) for the small-cap funds. Alphas are presented for the total sample period and each of the five sub-periods. The figure uniformly documents clear improvement in performance of small cap funds once the underperformance of FTSE Small Cap index as their benchmark is taken into account through AGT adjustment. The greatest performance enhancement over standard FF3 alpha is in the bear period 2000-2002, when the index was at its lowest: the small cap funds alpha increases from -0.48% to 7.1%. This supports our earlier findings that greater improvement in alphas is expected during market downturns.

Figure 4: Equity mutual fund (before and after AGT- adjustment) and FTSE Small Cap Index alphas

Figure shows for different time periods the FTSE Small Cap index alpha (FF3 model), the alpha of the Small cap sub-sample of equity mutual funds before and after adjustment the non-zero benchmark index alpha. This is done for the full time period (1992-2013), January, 1992 to December 1999, January 2003 to December 2007 and January 2010 to October 2013 (bull market) and January 2000 to December 2002 and January 2008 to December 2009 (bear market).



In summary, substituting a more general UK market index, FTSE 100, with a style specific benchmark, FTSE Small Cap, in the AGT model for a sample of UK small cap funds does not change our findings; it reinforces them and confirms that UK equity mutual fund performance

[19] As the results for Carhart alphas are qualitatively the same, we do not report them due to space considerations

is better than what the prior literature suggests. We believe these results will remain robust to the choice of other UK style specific indices as benchmarks, as they are highly correlated to FTSE 100. For instance, the correlations between FTSE UK Value Index and FTSE UK Growth Index with FTSE 100 are 0.95 each; while FTSE 250 Index that serves as a proxy for mid-cap companies has correlation of 0.84 with FTSE 100[20].

3.5. Conclusion

We take a new look at the performance of UK active equity mutual funds in light of recent academic evidence which suggest that indices funds select as benchmarks contain alphas. Therefore, the inferences one makes about the stock picking skills of fund managers stemming from standard performance measurement models such as Fama-French and Carhart, may be wrong as they embed benchmark alphas. In this study, we follow Angelidis et al. (2013) approach that suggests the use of benchmark adjusted alphas to shed a new light on performance measurement. Ours is the first study to document these benchmark-adjusted alphas for a sample of UK equity mutual funds. Our sample comprises of 887 active funds in the period of January 1992 to October 2013.

In our preliminary analysis, similar to some studies conducted on the US market such as Cremers et al. (2012), we report non-zero alphas of a passive benchmark index FTSE 100, used as a benchmark for all the funds in this study. However, in contrast to the US evidence, our findings indicate a significant negative benchmark index alpha of -1.12% for the Fama and French three-factor model and the annual alpha of -1.13% for the Carhart four-factor model, both statistically significant at 1% level. In addition, we show that benchmark index alphas vary in accordance to different market conditions; being significantly larger in bear market (between -1.61 and -2.86%) then bull market (-0.47 and -1.10%).

Most importantly, we reveal that both the Fama and French three-factor and Carhart four-factor models amplify the underperformance of UK equity mutual funds. After the Angelidis et al. (2013) adjustment for the negative alphas in the benchmark index, we show that UK focused equity funds are able to deliver positive excess performance, which is better than previous UK evidence suggests. As an illustration, for the whole sample period and the whole sample of funds, the Fama-French alpha exhibits ten-fold increase from just 13.81bps to 143.64bps per

[20] Source of data for all indices mentioned (inclusive of dividends): Datastream

year when adjusted for the negative alpha in FTSE 100. The adjustment brings greater increase in alphas in bear rather than in bull market periods, as the benchmark index performance was more depressed during market downturns. For instance, the financial crisis period of 2008-2009 bares the adjusted Fama-French annual alpha which is 2.89% higher than standard alpha for the sample of our funds. These results fare well with Angelidis et al. (2013), who show improvement in US mutual fund alphas after adjusting them for the funds' self-reported benchmarks.

Further, to test if the findings are consistent across funds' investment styles, we split the funds into nine style categories given by Morningstar style-box. When adjusted, alphas in all fund categories improve: when their value given by the standard Fama-French-Carhart models was negative, they became less negative (even positive, albeit mostly insignificant); when the standard alphas were positive, the AGT-adjustment brought them to a higher positive and significant level. We also find that over 70% of mutual funds concentrate their portfolios in Small/Value, Small/Growth and Small/Blend stocks. They perform better than other styles (generating positive AGT-adjusted FF3 alpha of 1.62%, 2.04% and 1.54%, respectively; statistically significant at 1% level). In these style groups, positive abnormal performance persists even during market downturns. Small/value style exhibits the most consistent outperformance, small/growth performance is driven largely by the dot.com boom, while large/value funds do better than any other group during the financial crisis 2008-2009. We conduct a robustness test for the choice of benchmark index. We find that replacing FTSE 100 benchmark with style-specific FTSE Small Cap Index for small cap funds in our sample strengthens and corroborates our results.

Overall, our study shows that adjusting fund alphas, obtained from standard Fama-French-Carhart performance measurement models, by the alpha of the benchmark shows improvement in UK equity mutual fund performance. Specifically, conclusions from previous empirical studies based on standard performance measures strongly tilt towards significant underperformance of UK funds. We show opposing evidence from AGT-adjusted alphas, in support of significant outperformance of UK equity funds, even during bear market periods. The study could be extended to the assessment of the conditional vs unconditional adjusted alphas and market timing ability of funds as in Ferson and Warther (1996) or the new look at the persistence in performance, similar to Fletcher and Forbes (2002).

Chapter 4
Second empirical essay

Abstract

This study re-visits the question of benchmark mismatch among 1281 US equity mutual funds and its impact on benchmark-adjusted fund performance and ranking. All funds report S&P500 index as a prospectus benchmark, yet 2/3 of those are placed in the Morningstar category with risk and objectives different to those of the S&P500 index. We identify 'true' benchmarks for those mismatched funds and find that their S&P adjusted alphas are higher than 'true' benchmark adjusted alphas in 61.2% of the cases. In terms of fund quartile rankings, 30% of winner funds lose that status when the prospectus benchmark is substituted with a more suited one. In the remaining performance quartiles there is no clear advantage of using S&P 500 as a prospectus benchmark. The prospectus benchmark therefore can mislead investors about fund's relative performance. This leads us to conclude that any reference to performance in a fund's prospectus should be treated with caution.

Keywords: Prospectus benchmark selection, Mutual fund benchmark mismatch, Benchmark-adjusted alphas, Performance ranking

JEL classification:**:** G11, G12, G23

4.1. Introduction

SEC regulations require mutual fund companies to disclose their performance relative to a passive benchmark, an index often referred to as their prospectus benchmark. Over a third of US investors rely on information in the fund prospectus when purchasing a mutual fund[21]. Prospectus benchmark defines an investment direction and a risk tolerance, and should reflect the strategic role of the individual asset classes in the fund. However, Cremers and Petajisto (2009) provide evidence that mutual funds typically have a high proportion of holdings that differ from those of fund's (theoretically adequate) benchmark index. Sensoy (2009) affirms that funds frequently differ significantly from their benchmarks and shows that value funds are more likely to have self-designated benchmarks that are mismatched on value/growth, while small-cap funds tend to have prospectus benchmarks mismatched on size.

It should not come as a surprise then that some prospectus benchmark choices may be misleading, as there are no precise requirements on the selection of funds' best suited benchmark. Therefore, the choice of fund benchmark may be biased and may indicate principal-agent problems. As a consequence, for instance, a fund reporting a large cap index as their prospectus benchmark may have significant proportion of their assets invested in smaller size stocks. Considering investors' close scrutiny of fund performance it is vital to examine the extent of benchmark misclassification in US active fund management. Moreover, considering the development of recent literature on mutual fund performance, it is crucial to account for non-zero benchmark alphas, which significantly bias outcomes of fund performance (see for instance Chinthalapati et al., 2017). A recent study by Cremers, Petajisto and Zitzewitz (2012) shows that standard benchmark models produce economically and statistically significant non-zero alphas for passive benchmark indices, including a widely used US passive benchmark - the S&P 500. Negative and statistically significant alpha for the Russell 2000 Growth index was documented by Chan, Dimmock, and Lakonishok (2009); significant non- zero alphas are also discussed in Costa and Jakob (2006).

[21] Investment Company Institute, Understanding Investor Preferences for Mutual Fund Information, Summary of Research Findings ("Understanding Investor Preferences"), 2006, available at https://www.ici.org/pdf/rpt_06_inv_prefs_full.pdf

Based on the above, this paper aims to examine to which extent the benchmark choice of US long only equity funds changes inferences on fund performance, once the benchmark alphas are accounted for in fund performance evaluation. In particular, we assess whether inadequate prospectus benchmark selection may lead to over estimation of fund performance and whether it could be a subject of gaming. Further, we investigate whether benchmark choice affects fund performance in relative terms (relative to peers) and, therefore, changes the ranking position of the winning and losing funds, in particular. Hence, as our main contribution, we add to the literature on US mutual fund benchmark mismatch by 1) investigating the impact of the choice of benchmark on fund performance and performance rankings and 2) providing performance assessment free of biases caused by alphas embedded in the benchmark index and not accounted for in the standard pricing models. To account for these non-zero benchmark alphas, we apply Angelidis, Giamouridis and Tessaromatis (2013) methodology that allows for the alpha in the benchmark index to be included in a standard factor model, such as Carhart (1997). This approach adjusts alpha of a fund by that of the benchmark.

In the aspect of previous literature relevant to analysis, Sensoy (2009) provides evidence that funds frequently differ from their benchmarks in terms of their risk characteristics and composition for strategic reasons. Substantial exposures to size and value/growth factors in returns that are not captured by their benchmarks were also discussed in Elton, Gruber, and Blake (2003). The study of DiBartolomeo and Witkowski (1997) examine monthly returns for 748 load and no-load open-end funds and show that return patterns of 40 percent of funds analysed deviate from the benchmark declared in the prospectus with 9 percent of funds being seriously misclassified, two or more risk tiers away from their declared categories. Similarly, Kim, Shukla and Tomas (2000) assess how well mutual funds' stated objectives conform to their attributes-based objectives and revealed that the stated objectives of more than half the 1043 funds analysed differ from their attributes-based objectives, and over one third of the funds are severely misclassified. The study also confirms upward and downward risk shifts. Bams, Otten, and Ramezanifar (2016) analyse a sample of 1,866 US equity funds over the 2003-2015 period and found that 14% of funds are significantly misclassified based on long term style analysis. Huang et al. (2011) show that mutual funds change their total risk exposure substantially over time. Authors claim that it might be done for strategical reasons: in order to increase the expected money inflows to the funds or to manipulate their performance numbers. Similarly, Kacperczyk, Sialm, and Zheng (2008) measure the return gap, the difference between the reported fund return and the return on a portfolio that invests in the previously disclosed fund holdings, and document that despite disclosure requirements, mutual fund investors do not observe all actions of fund managers. Portfolio performance manipulation and deviation from benchmarks was also discussed in Goetzmann et al. (2007), Jiang et al. (2014), Fung and Hsieh (2002).

This paper contributes to the mutual fund performance measurement literature. In addition it adds to the literature on mutual fund benchmark misclassification and extends the work of Chan, Dimmock, and Lakonishok (2009), which demonstrates that judgments about the magnitude of performance are sensitive to the benchmarking methodology. To the best of our knowledge, this is the first study that analyses the impact of benchmark choice on US equity fund performance and ranking while accounting for the non-zero alpha bias in those passive benchmarks. We use the net monthly returns of 1281 actively managed US equity mutual funds from January 1992 to February 2016. All funds in the sample report S&P500 as their primary prospectus benchmark in the Morningstar database. Our funds belong to 24 distinct Morningstar global categories: e.g. US Small Cap, US Large Cap Value, Energy Sector Equity, Global Equity etc. Investigation of commonly used benchmarks amongst funds in different categories in the Morningstar database, shows us that the primary prospectus benchmark that all our funds use, the S&P 500 Index, is most suitable for the funds in the Large Cap Blend Morningstar category. However, around 2/3 of the funds in our sample are not in that category, yet they declare S&P 500 as their passive benchmark. Our analysis of prospectus benchmark fit shows that the funds' rationale for selecting a particular passive index as prospectus benchmark is not clear, as the index does not correspond to funds composition or investment objectives in large proportion of our sample. For each of the Morningstar global categories, we identify a more appropriate benchmark than the S&P 500, which we refer to as 'true' benchmark in this paper. We find that 'true' benchmarks are a better fit for our funds than their prospectus benchmark, the S&P 500 index, having on average around 10% higher R-squared in the full sample period and each of the sub-periods. This makes an inference that even adjusted for benchmark alphas fund performance may be significantly biased if fund performance estimated versus unsuitable/prospectus benchmark is used by investors as a performance target.

To measure fund performance and rank the funds we apply Angelidis et al. (2013) methodology (AGT hereafter) that adjusts fund's alpha for benchmark's alpha, hence isolating manager's skill above that common to the benchmark. We find that 61.2% of the mutual fund AGT alphas are higher when S&P500 is used as a benchmark[22]. Further, in 15 out of 22 rolling periods of 36 months each, pairing the performance with S&P500 is beneficial to the funds and leads to overestimated performance. Thus, on average, prospectus benchmark amplifies fund performance by 23 basis points versus the performance adjusted with a 'true' benchmark. Nevertheless, there is still the remaining 30 percent of periods when performance is better when the 'true' benchmark alpha is used as the target in AGT model.

Analysis of fund quartile rankings shows that, on average, around 30% of winners leave the top quartile of funds when the benchmark is changed from the self-designated benchmark, S&P 500, to the 'true' benchmark in AGT benchmark-adjusted alpha estimation. On the opposite end of spectrum, nearly 30% of losers move up the quartiles when the 'true' benchmark is used. This shows that our results support the notion from Sensoy (2009) that the funds that appear at the top end of the spectrum may choose their prospectus benchmarks strategically. However, inappropriately chosen prospectus benchmark actually harms the funds that are at the bottom of the ranks. Given this, we conclude that the choice of the appropriate benchmark is critically important, as the wrong benchmark does not only bias performance assessment but can also lead to false conclusions when performance of funds relative to peers is assessed. Hence, this paper is of significant importance to individual investors, institutional investors and professional financial advisors interested in performance evaluation and fund rankings. Moreover, it has implications for financial regulators and policy makers with respect to fund information disclosure requirements and transparency in benchmark selection.

This paper proceeds as follows: Section 2 describes the data. Section 3 provides preliminary analysis where we test the existence of benchmarks alphas and check which benchmarks has a better explanatory power to the fund investment style. Section 4 presents Methodology. Section 5 analyses funds' AGT-adjusted alpha performance and provides results. Section 6 delivers outcomes on biases in quartile ranking. Section 7 concludes.

[22] The results presented are obtained with the use of the Carhart model in AGT augmentation. The outcomes obtained with Fama-French three and five factor models are qualitatively the same and available upon request.

4.2. Data

The data set is comprised of 1,281 long-only active US equity mutual funds from January 1992 to February 2016. The net monthly returns of mutual funds are from Morningstar, inclusive of dividends. All funds have minimum requirement of 36 months of returns to be included in the sample. There is no survivorship bias in the sample. All funds in the sample declare S&P500 as their prospectus benchmark. However, due to the fact that in some cases investment objective stated in a fund's prospectus may not reflect how the fund actually invests (Kacperczyk, Sialm, and Zheng, 2008, Sensoy, 2009, Huang et al. 2011) Morningstar offers own proprietary data where each fund is assigned to a Morningstar Global Category based on the underlying holdings and fund's portfolio statistics. The list of Global Categories Morningstar has assigned our funds to, the number of funds per each category, the most relevant passive benchmark[23] for each category and the number of monthly observations per category are presented in Table 1. The returns data for all benchmarks is inclusive of dividends.

Table 1: Sample of 'true' benchmarks

The sample consists of 1,281 (212,122 monthly observations) long-only active US equity mutual funds from January 1992 to February 2016. For all funds the self-declared prospectus benchmark is the S&P500. Table below shows the Morningstar Global Category our funds belong to, the suitable benchmark for the category, the number of funds in the category and number of monthly observations per category (all benchmarks are total return and in USD).

Global Category	Suitable Benchmark	# Funds	# Monthly Observations
US Large Cap Blend	S&P 500	460	73,493
US Large Cap Growth	RUSSELL 1000 GROWTH	290	48,393
US Large Cap Value	RUSSELL 1000 VALUE	127	21,160
US Mid Cap	RUSSELL MIDCAP	112	17,332
Technology Sector Equity	S&P500 ES INFO TECHNOLOGY	54	9,092
US Small Cap	RUSSELL 2000	40	5,611
Healthcare Equity	S&P500 ES HEALTH CARE	32	5,554
Real Estate Equity	S&P500 DIVERSIFIED REIT'S	24	2,279
Global Equity	MSCI WORLD	22	3,392
Financial Sectors Equity	S&P500 DIVERSIFIED FINANCIALS	19	4,162
Energy Sector	S&P500 ENERGY IG	16	3,198
Precious Metals Sector Equity	S&P GSCI Precious Metal Tot. Ret.	16	4,196
Utilities Sector	S&P500 ES UTILITIES	14	3,293
Natural Resources Equity	S&P GSSI NORTH AMER. NAT.RES.SECTOR	13	2,400

[23] We review the indices that all available US equity funds in a given Global category benchmark against and select the most common benchmark, ensuring that its characteristics correspond to the category it represents (US Large cap value is best represented by Russell 1000 value Index etc.)

Consumer Goods and Services	S&P500 ES CONSUMER DISCRETIONARY	12	3,159
Industrials Equity	S&P500 ES INDUSTRIALS	8	2,124
Communications Equity	S&P500 COMM. EQUIPMENT	8	1,268
Global Equity Large Cap	MSCI EAFE	7	1,174
Emerging Markets Equity	MSCI EM	2	324
Other Equity (Emerging Europe)	MSCI EM EUROPE	1	227
Europe Large Cap Equity	MSCI EUROPE	1	82
Asia Equity ex Japan	MSCI AC ASIA PAC EX JP	1	58
Japan Equity	MSCI JAPAN	1	61
Greater China	MSCI GOLDEN DRAGON	1	90
	Total:	**1,281**	**212,122**

Only 36% of our sample (460 funds) fall in the Large Cap Blend Morningstar Global category where the S&P 500 would be deemed as the most appropriate passive benchmark. It means that performance analysis where the fund performance is measured against a prospectus benchmark can be biased and can provide inaccurate inferences about manager's skill. Further, 32% of our funds belong to the Large Cap Value and Large Cap Growth Global category where most commonly used benchmark for funds are Russel 1000 value and Russell 1000 Growth index respectively. Midcap Global Category is also highly presented (112 out of 1,281 funds, 8.7%). Its better fit would be Russell Midcap Index rather than S&P 500. Some of our funds are in the Small Cap Morningstar Global category (40 of 1,281), best represented by a Russell 2000 index. Overall, these aforementioned five categories (whose appropriate benchmarks should be S&P 500, Russel 1000 Value, Russel 1000 Growth, Russel Midcap and Russell 2000 index) account 80 percent of our sample. All other Morningstar Global categories in our sample are sector specific or country/region specific and call for sector or regional benchmarks. These specialist funds account for the remaining 20% of our sample, Hence, significant proportion (64%) of our funds selects and reports a benchmark inappropriate for their category of funds. This is important from two perspectives: 1) measuring fund performance relative to the benchmark and 2) measuring fund performance relative to other similar funds. To this end, it is important to investigate fund's relative rankings within the same category and assess whether the funds that are the top performers according to prospectus benchmark (S&P500) change their relative ranking position after their performance is calculated with a benchmark that better reflects the risk characteristics of their Morningstar Global Category. Section 3.1 provides further discussion on suitability of the funds' self-declared benchmarks.

We split our analysis in 22 rolling overlapping sub-periods, each being 36 months of length. Given that the minimum data requirement for each fund is 36 months, within each rolling period we require that a fund has no less than 30 months of continuous returns. Table 2 reports the number of funds and monthly observations for each of the rolling sub-periods:

Table 2: Sample funds with more than 36 monthly observations

Table reports the number of funds and monthly observations for each of the 36 months rolling windows. The minimum data requirement is for funds to have at least 36 months of continuous observations and no less than 30 months of continuous returns within each rolling period. The #Funds represents the number of (non-unique) funds with available data in each period.

Period	*# Funds*	*# Monthly Observations*	*Period*	*# Funds*	*# Monthly Observations*
199201:199412	409	12,508	*200301:200512*	1,034	32,887
199301:199512	451	14,042	*200401:200612*	1,070	33,361
199401:199612	527	15,740	*200501:200712*	1,066	33,956
199501:199712	600	17,860	*200601:200812*	1,054	34,366
199601:199812	681	20,463	*200701:200912*	1,057	33,663
199701:199912	771	23,364	*200801:201012*	1,039	32,453
199801:200012	865	26,305	*200901:201112*	975	30,906
199901:200112	919	28,916	*201001:201212*	895	29,500
200001:200212	955	30,874	*201101:201312*	855	27,929
200101:200312	980	32,085	*201201:201412*	789	26,519
200201:200412	997	32,640	*201301:201602*	751	26,573
			Overall: 199201:201602	**1,281**	**211,855**

4.3. Preliminary analysis

4.3.1. Test on the appropriateness of benchmark allocation

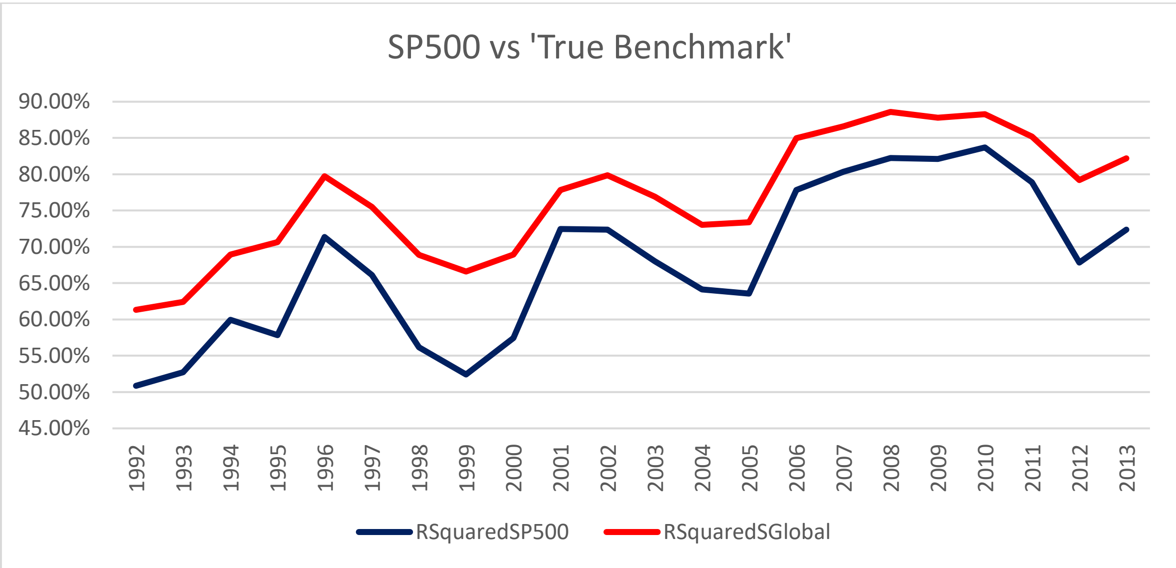
To begin with, we examine whether the 'true' benchmarks (the ones more appropriate for funds' Morningstar Global Category) provide a better fit than the self-declared prospectus benchmark. To estimate this, we use the R-squared from equations (1) and (2) as a proxy for the accuracy of the benchmark used:

$$R_{i,t} - R_{f,t} = \alpha_{i,} + \beta_{i,t} R_{S\&P500} - R_{\mathrm{F},t,} + e_{i,t} \quad (1)$$

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,t} R_{\text{'True' Benchmark}} - R_{\mathrm{F},t\cdot,} + e_{i,t} \quad (2)$$

In this analysis we exclude the mutual funds that belong to the Large Cap blend category (460 funds, as per Table 1) for which the 'true' benchmark (S&P500) is the same as their prospectus benchmark. For the remaining 821 funds in the sample we estimate equation (1) and (2), over the 22 rolling windows. In other words, we regress excess return of each out of 821 mutual funds on the excess return of the benchmark (SP500 and the most relevant Global category benchmark). Thus, we run regressions (for each fund and each out of 22 rolling periods) with equation (1) and obtain R-squared from each regression. A similar procedure is performed with equation 2. Figure 1 depicts the average R-squared across the funds per each sub-period estimated using the S&P 500 as the benchmark (lower line) and ‘true’ benchmarks/Global category benchmark (upper line), as per equations (1) and (2).

Figure 1: Average R-squared of S&P500 and ‘True’ benchmark fit



The results confirm our notion that ‘true’ benchmarks are more suited for funds outside the Large Cap Blend category than the S&P500 index. The R-squared obtained using ‘true’ benchmarks for each sub-period and for the entire sample period is on average 10% higher, with peaks in 1999 and 2012, when the difference reached 14% and 11.5%. Given these results, the question that imposes itself is that of the impact of poorly suited benchmarks on the mutual fund performance and their ranking relative to other funds: do funds with a prospectus benchmark unsuitable for their investment style tend to outperform those benchmarks and whether they remain at the top of the fund rankings when the benchmark is swapped for the ‘true’ one.

4.3.2. Presence of alphas in passive benchmarks

The second issue we try to avoid in our assessment of performance ranks is that of the 'closet-indexing'. For instance, if a 'true' benchmark (say, Russell 1000 Value) performs better than the self-designated benchmark chosen by a fund (S&P 500 here), the fund that belongs to that specific category (Large Cap Value in this example) is likely to outperform its self-reported benchmark (S&P 500), even if they are simply replicating their 'true' benchmark (Russell 1000 Value). That also means that such funds may rank higher relative to some other funds even though the fund managers exhibit no true skill.

To illustrate such bias inflicted by indices, in spirit of Costa and Jakob (2006), Chan, Dimmock, and Lakonishok (2009), Cremers, Petajisto and Zitzewitz (2012)[24], we estimate standard Carhart four-factor alphas of both self-declared prospectus benchmark (S&P500) and the 'true' benchmarks in our sample:

$$R_{Benchmark,t} - R_{f,t} = \alpha_{Benchmark} + \beta_{M,t}\left(R_{M,t} - R_{F,t}\right) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{WML}WML_t + e_t \qquad (3)$$

Where $R_{Benchmark,t}$ is the return on the (prospectus or 'true') benchmark index used; R_f is the US 1 month Treasury bill; $R_{M,t} - R_{F,t}$ is the market risk premium[25]; SMB and HML are size ad value factors from Fama and French (1993) paper and WML is the Carhart (1997) momentum factor. $\alpha_{Benchmark}$ represents the four-factor (prospectus or 'true') benchmark alpha, i.e. the excess return of the benchmark unexplained by the four factors, e_t is error term.

The four-factor Carhart alpha is calculated for the S&P500, Russel 1000 Growth, Russel 1000 Value, Russel Midcap and Russell 2000 over 36 monthly rolling periods, to obtain alphas from 1994 to 2016. The aforementioned benchmarks correspond to the five largest Morningstar Global categories in the data set and represent 80 percent of our fund sample (1,029 funds of a total 1,281). The remaining indices and their corresponding categories in our sample are not used for this analysis as the number of funds per category is not large enough resulting in some sub-periods featuring very few funds, jeopardising the objectivity of the results.

[24] who report non-zero alphas for passive benchmark indices

[25] US market risk premium is defined as the value-weighted return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ (R_m) minus one month US Treasury bill (R_f)

Figure 2: Four factor (Carhart) alphas of S&P 500 and selected 'true' benchmarks

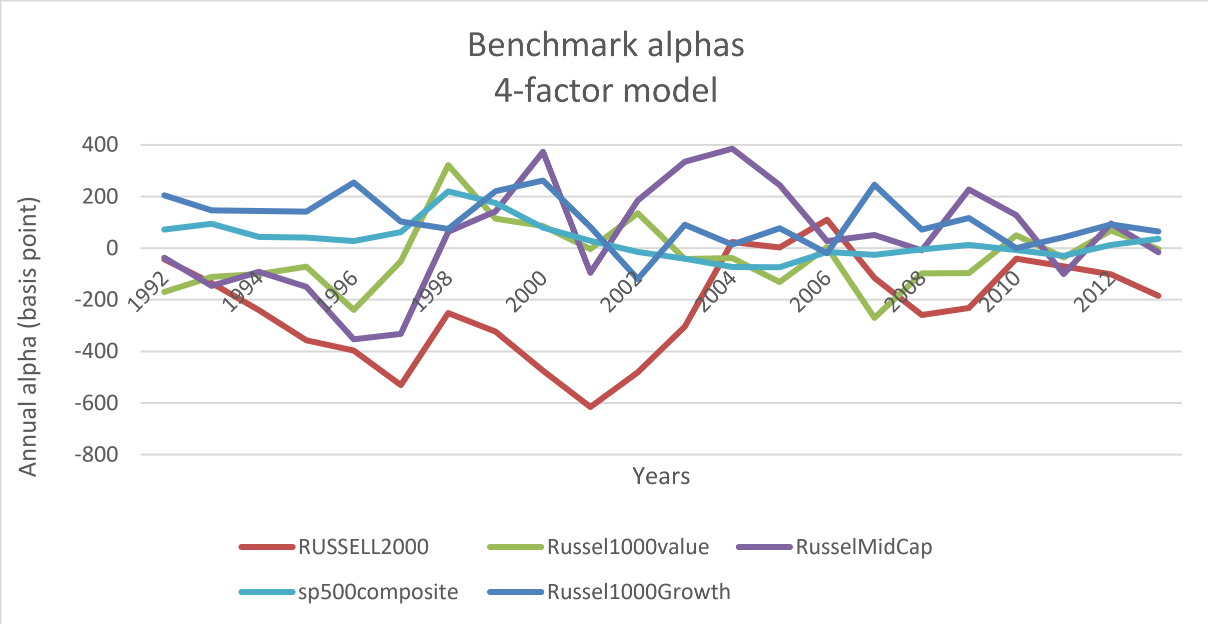


Figure 2 depicts the trend of annualized four-factor alphas (in bps) of the five indices. First, in line with previous studies (see for instance Cremers, Petajisto and Zitzewitz (2012), Chinthalapati et al., 2017) the alphas of the five passive benchmarks are not zero. Specifically, the S&P500 and Russell 1000 Growth alphas more stable and tend to be more positive than those of the remaining indices analysed here. In the full sample period from January 1992 to February 2016, the S&P 500, Russell 1000 Growth and Russell Midcap indices all have positive four-factor alphas of 33.01, 74.93 and 60.17 basis points per year respectively; while the negative alphas of -12.58 and -197.01 basis points per year are obtained for the Russell 1000 Value and Russell 2000 index. Qualitatively similar results were obtained when the Carhart model was substituted by the Fama and French three and five-factor models.

To obtain an indication of the magnitude of possible biases in fund performance evaluation by selecting an index not corresponding to funds' risk profile and composition holdings, we calculate the difference between the Carhart alpha of the 'true' benchmark and the self-declared benchmark, S&P500, as per Figure 3. The difference is annualized and reported in basis points.

Figure 3 Differences between Carhart alphas of selected 'true' benchmarks and the S&P500

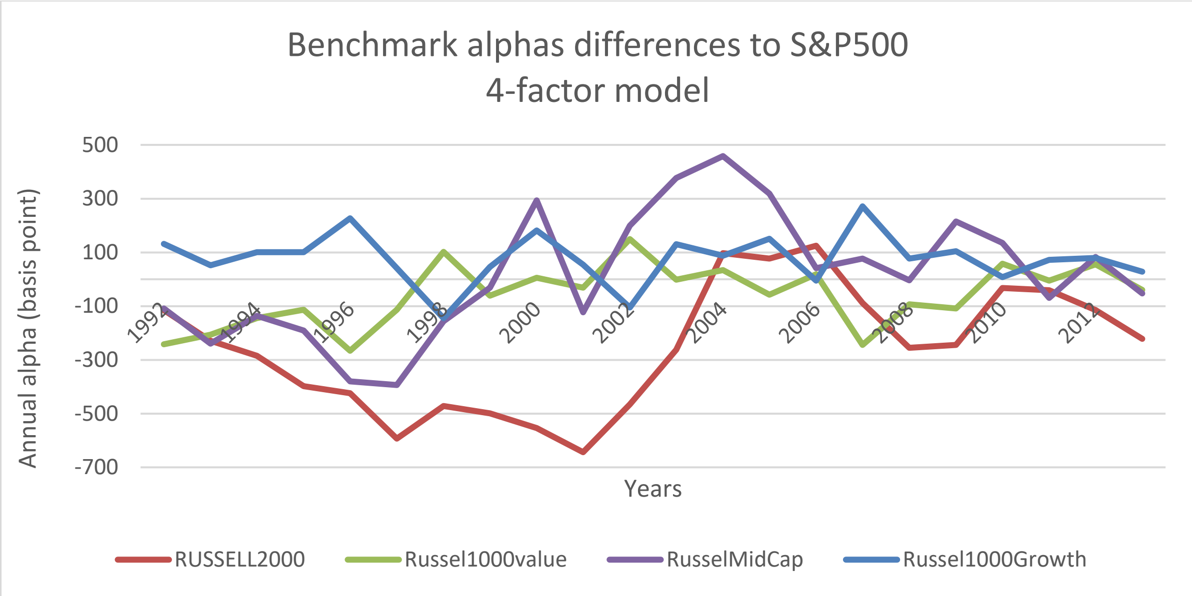


Figure 3 illustrates that S&P500 four-factor alphas differ from the remaining four indices corresponding to the global categories where most of our funds 'reside'. For instance, in the sub-periods 1994-1996 and 1996-1998, the alpha for the self-declared prospectus benchmark S&P500 is positive but at least 100bps lower than the alpha for Russel 1000 Growth index. This tendency of the 'true' benchmark alpha to be higher than the prospectus benchmark one is present in 20 out of 22 rolling windows in this study. Therefore, a mutual fund which is in Large Cap Growth category may take the benefit of the lower prospectus benchmark alpha relative to the 'true' one, more typical for its composition and risk. If such fund "beats" the prospectus benchmark, investors may view that as a vouch for managerial skill, whereas the fund may be simply replicating Russell 1000 Growth, not having any stock picking skill. Hence, its outperformance over prospectus benchmark should simply be attributed to a higher alpha of the 'true' benchmark, more appropriate for the given fund. However this is not the case for all the indices. Inverse situation can be noticed for Russel 2000, whose four-factor alpha is systematically lower than the S&P500 one.

To avoid the impact of these 'closet indexers', there is a need to look at the benchmark-adjusted performance of funds. This is particularly important to note when measuring funds 'true' performance and 'true' ranking relative to other funds both within the Morningstar Global

Category peer group and overall. In the following section we present the methodology that appropriately adjusts performance and provides funds' benchmark-adjusted alphas.

4.4. Performance and ranking methodology

The Fama and French (1993) three-factor model and the Carhart (1997) four-factor model, utilised in this paper, are standard models, most widely known and accepted in the industry for assessing portfolio alpha-generating ability

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{M,i}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + e_{\mathrm{i},t} \quad (4)$$

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{M,i}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{WML,i}WML_t + e_{i,t} \quad (5)$$

where $R_{i,t}$ is the return of a mutual fund *i* in period *t*, α_i is the excess return of the fund *i* over period *t*, R_{Ft}, is risk free rate, R_{Mt} is the total monthly return (inclusive of dividends) on the value-weigh market portfolio, *SMB* and *HML* are Fama and French (1993) size (small minus big returns) and value (high minus low book-to-market returns) factors respectively, *WML* is Carhart (1997) the momentum (winner minus loser returns) factor, $e_{i,t}$ is error term.

To stretch our analysis even further we also test equity fund performance using the last Fama and French five factor model.

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{M,i}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{RMW,i}RMW_t + \beta_{CMA,i}CMA_t + e_{i,t} \quad (6)$$

where *RMW* and *CMA* are Fama and French (2016) profitability (robust minus weak) and investment (conservative minus aggressive) factors respectively, and the rest of the variables are described as per equations (5) and (6).

To obtain unbiased alphas for funds, we apply Angelidis, Giamouridis and Tessaromatis (2013) adjustment, suggested in recent literature on performance measurement. The model is of interest to academics and investment professionals, as it adjusts the left hand side of the standard Carhart (1997) as well as Fama and French three and five-factor models by replacing the risk-adjusted return with the benchmark-adjusted return, i.e. the funds' tracking error:

$$R_{i,t} - R_{Benchmark,t} = \alpha_i^* + \beta_{i1}^*\left(R_{M,t} - R_{F,t}\right)_t + \beta_{i2}^*SMB_t + \beta_{i3}^*HML_t + e_{i,t}^* \quad (7)$$

$$R_{i,t} - R_{Benchmark,t} = \alpha_i^* + \beta_{i1}^*(R_{M,t} - R_{F,t})_t + \beta_{i2}^* SMB_t + \beta_{i3}^* HML_t + \beta_{i4}^* WML_t + e_{i,t}^* \quad (8)$$

$$R_{i,t} - R_{Benchmark,t} = \alpha_i^* + \beta_{i1}^*(R_{M,t} - R_{F,t})_t + \beta_{i2}^* SMB_t + \beta_{i3}^* HML_t + \beta_{i4}^* RMW_t + \beta_{i5}^* CMA_t + e_{i,t}^* \quad (9)$$

where $R_{i,t} - R_{Benchmark,t}$ is the excess return of a mutual fund *i* over a benchmark in period *t*. As in equation (3) SMB and HML are size ad value factors from Fama and French (1993) paper, WML is the Carhart (1997) momentum factor, *RMW* and *CMA* are Fama and French (2016) profitability (robust minus weak) and investment (conservative minus aggressive) factors respectively. α_i^* is the difference of the fund's and benchmark's FF three, five and the Carhart alpha (AGT-adjusted alpha hereafter). Similarly, Beta ($\beta_{i1}^*, \beta_{i2,}^* \beta_{i3,}^* \beta_{i4}^*, \beta_{i5}^*$) is the difference between the fund's and benchmark's FF three, five and the Carhart betas. All of the factor data is from Kenneth French's website[26].

The AGT model, therefore, enables us to obtain AGT-adjusted alpha of a fund that accounts for the alpha of the benchmark. To assess the change in rankings when the benchmark changes from the prospectus benchmark to the 'true' category benchmark, each of the three aforementioned models will be used twice for each equity fund: with the S&P 500 as a benchmark and with the 'true' benchmark relevant for the Morningstar category a fund belongs to:

As an example, AGT-adjusted four-factor alphas will be estimated as follows:

$$R_{i,t} - R_{S\&P500,t} = \alpha_i^{*S\&P500} + \beta_{i1}^{*S\&P500}(R_{M,t} - R_{F,t})_t + \beta_{i2}^{*S\&P500} SMB_t + \beta_{i3}^{*S\&P500} HML_t + \beta_{i4}^{*S\&P500} WML_t + e_{i,t}^{*S\&P500}. \quad (10)$$

$$R_{i,t} - R_{True,t} = \alpha_i^{*True} + \beta_{i1}^{*True}(R_{M,t} - R_{F,t})_t + \beta_{i2}^{*True} SMB_t + \beta_{i3}^{*True} HML_t + \beta_{i4}^{*True} WML_t + e_{i,t}^* \quad (11)$$

Where $R_{S\&P500,t}$ and $R_{True,t}$ are the return of the S&P 500 and 'true' Morningstar Global category benchmark respectively, $\alpha_i^{*S\&P500}$ is the difference between the Carhart alpha of fund *i* and its prospectus benchmark, S&P 500, α_i^{*True} is the difference between the Carhart alpha of fund I and the 'true' benchmark; $\beta_{i1}^{*S\&P500}, \beta_{i2,}^{*S\&P500} \beta_{i3,}^{*S\&P500} \beta_{i4}^{*S\&P500}$ are fund *i* exposures to market risk, size, style and momentum factors beyond the exposure of S&P500 to those risks and $\beta_{i1}^{*True}, \beta_{i2,}^{*True} \beta_{i3,}^{*True} \beta_{i4}^{*True}$ are the fund *i's* four-factor betas adjusted by those of the 'true' benchmark for fund *i*'s category. The rest is as per equation (8).

[26] Kenneth French's website: http://mba.tuck.dartmouth.edu/pages/faculty/ ken.french/data_library.html.

4.5. Performance Results

Preliminary to the main assessment we perform the Hausman test to determine which model, fixed-effects or random-effects, is more appropriate to the sample data. Based on the results we continue with a fixed effects panel model estimation procedure.

To estimate the impact of benchmark choice on fund performance we calculate Fama and French (1993, 2016) and Carhart (1997) AGT-adjusted alphas for all funds in each 36 months rolling window. We run double regression against S&P500 and 'true' category benchmark for each of the 22 rolling overlapping sub-periods. Even though the cut off is to include all funds with at least 36 months of continuous observations during the whole period; less than 36 months can be observed per each sub-period analysed (see Table 2). Within each rolling period we require that a fund has no less than 30 months of continuous returns. Table 3 lays out the results. The indicated 36 month rolling window is a "typical" investment horizon applied by professionals and referred in academic literature (see for instance Brown and Goetzmann, 1997; Ben-Rephael et al., 2012; Petajisto 2013). Overlapping assumes that investors re-access their investments in a year basis. Overall, results estimated with overlapping and non-overlapping periods are qualitatively similar.[27]

As it can be seen from the outcomes, at this stage it is not obvious that mutual funds take advantage of reporting S&P500 as their prospectus benchmark. The results are mixed. There are time periods such as 1997-1999 and 2002-2004 when alphas estimated with S&P500 are negative and lower comparing to the ones from 'true' category benchmark (statistically significant at 1 percent level for the three models), so that means that benchmarking against S&P 500 allows funds overestimate their performance. In opposite, for the years 2011-2013, 2012-2014 the tendency is different, higher negative and significant alphas deteriorate the funds' performance when regressed on S&P500. There are also the cases when the performance differs depending on the models applied (in line with Chan, Dimmock, and Lakonishok, 2009). Thus, higher negative and significant alphas are recorded for FF3 and Carhart models for the years 1994-1996, 1996-1998, 2001-2003 but the performance is better

[27] Other studies such as Swinkels and Van Der Sluis (2006), Mamaysky et al., (2008) propose to use Kalman filter to estimate the dynamics of mutual fund alphas and betas. However, the aim of our paper is to look to alpha variation using unconditional standard pricing models and not beta variation (styles).

Table 3: Mutual fund alphas estimated with FF3, FF5 and the Carhart models for both S&P 500 and 'true' benchmarks

Panel A-B: Represents results for S&P 500 and 'true' benchmarks (calculated for all funds and the total sample period, and for all funds for each of each of the 22 rolling overlapping sub-periods (as in Table 2)) Each sub-period includes all funds with at least 36 months of continuous observations during the whole period, with a minimum of 30 months of continuous returns within each rolling period. Panels are employed with fixed effects (according to the Hausman test performed). Alphas are annualised and given in basis points.

Panel A: Alphas for S&P 500

Benchmark S&P500	***FF 3-factor***		***Carhart 4-factor***		***FF 5-factor***			
	Alpha bp	R-Squared	Alpha bp	R-Squared	Alpha bp	R-Squared	#Funds	#Obs
199201:201602	-74.96*** (-8.75)	0.0503	-81.80*** (-9.45)	0.0504	-19.09** (-2.12)	0.0542	1,281	211,855
199201:199412	1.00 (0.02)	0.1007	-61.73** (-2.05)	0.1081	163.88*** (4.87)	0.1052	409	12,508
199301:199512	-25.30 (-0.92)	0.0907	-82.55*** (-2.84)	0.0930	84.47*** (2.69)	0.0917	451	14,042
199401:199612	-143.43*** (-5.79)	0.1842	-140.25*** (-5.32)	0.1833	-66.80** (-1.93)	0.1223	527	15,740
199501:199712	-284.56*** (-8.76)	0.1627	-303.62*** (-9.04)	0.1630	34.63 (0.77)	0.1702	600	17,860
199601:199812	-196.14*** (-6.26)	0.1008	-211.23 (-6.51)	0.1009	-35.68 (-1.09)	0.1094	681	20,463
199701:199912	-215.12*** (-6.69)	0.0702	-186.37*** (-5.54)	0.0706	-273.65*** (-8.01)	0.0852	771	23,364
199801:200012	227.41*** (5.95)	0.0800	117.05*** (2.77)	0.0813	70.26 (1.59)	0.0921	865	26,305
199901:200112	104.05*** (2.65)	0.0595	67.52* (1.72)	0.0623	25.85 (0.61)	0.0750	919	28,916
200001:200212	-16.14 (-0.42)	0.0613	1.11 (0.03)	0.0646	177.32*** (4.09)	0.0670	955	30,874
200101:200312	-239.31*** (-8.47)	0.0325	-228.68*** (-8.10)	0.0350	-124.33*** (-3.88)	0.0321	980	32,085
200201:200412	-162.16*** (-7.44)	0.0569	-161.71*** (-7.42)	0.0573	-58.26** (-2.49)	0.0627	997	32,640
200301:200512	-23.12 (-1.31)	0.0637	-19.6 (-1.12)	0.0646	9.80 (0.54)	0.0700	1,034	32,887

200401:200612	-21.16 (-1.16)	0.0758	34.92* (1.89)	0.0840	-45.55** (-2.46)	0.0850	1,070	33,361
200501:200712	188.19*** (12.77)	0.0487	58.66*** (3.81)	0.0665	134.62*** (9.12)	0.0647	1,066	33,956
200601:200812	27.03 (1.42)	0.0354	-24.79 (-1.27)	0.0399	113.17*** (5.33)	0.0544	1,054	34,366
200701:200912	105.70*** (4.82)	0.0256	80.03*** (3.61)	0.0271	162.83*** (6.72)	0.0335	1,057	33,663
200801:201012	-12.07 (-0.51)	0.0194	-65.26*** (-2.74)	0.0246	140.47*** (5.52)	0.0315	1,039	32,453
200901:201112	-54.23*** (-2.66)	0.0263	-140.91*** (-6.83)	0.0387	29.63 (1.39)	0.0339	975	30,906
201001:201212	-220.87*** (-13.86)	0.0201	-187.30*** (-11.64)	0.0268	-43.20** (-2.42)	0.0377	895	29,500
201101:201312	-319.30*** (-18.96)	0.0277	-261.12*** (-14.99)	0.0338	-192.31*** (-10.25)	0.0389	855	27,929
201201:201412	-286.87*** (-15.19)	0.0251	-223.14*** (-11.36)	0.0305	-246.68*** (-12.65)	0.0439	789	26,519
201301:201602	-285.76*** (-14.76)	0.0268	-250.95*** (-12.61)	0.0290	-275.72 (-14.14)	0.0295	751	26,573

Panel B: Alphas for 'true' category benchmark

Benchmark Global Category	***FF 3-factor***		***Carhart 4-factor***		***FF 5-factor***			
	Alpha bp	R-Squared	Alpha bp	R-Squared	Alpha bp	R-Squared	#Funds	#Obs
199201:201602	-64.92***	0.0234	-83.75***	0.0228	-75.53***	0.0247	1,281	211,855
	(-6.97)		(-8.91)		(-7.75)			
199201:199412	-11.74	0.0669	-52.40*	0.0711	69.09**	0.0679	409	12,508
	(-0.45)		(-1.95)		(2.32)			
199301:199512	-45.41*	0.0622	-78.81***	0.0632	-1.00	0.0867	451	14,042
	(-1.84)		(-3.01)		(-0.04)			
199401:199612	-110.08***	0.1314	-122.90***	0.0728	-99.81***	0.1316	527	15,740
	(-4.70)		(-4.94)		(-3.06)			
199501:199712	-196.12***	0.1012	-215.86***	0.1015	-27.76	0.1037	600	17,860
	(-6.52)		(-6.93)		(-0.67)			
199601:199812	-180.51***	0.0484	-168.80***	0.0485	-91.08***	0.0503	681	20,463
	(-5.61)		(-5.07)		(-2.70)			
199701:199912	-251.33***	0.0537	-204.94***	0.0434	-275.81***	0.0682	771	23,364
	(-6.64)		(-5.17)		(-6.80)			
199801:200012	217.65***	0.0867	173.89***	0.0868	60.46	0.0872	865	26,305
	(4.27)		(3.08)		(1.02)			
199901:200112	122.00**	0.0185	69.19	0.0217	3.90	0.0237	919	28,916
	(2.27)		(1.28)		(0.07)			
200001:200212	-60.99	0.0254	-42.40	0.0280	-17.76	0.0285	955	30,874
	(-1.26)		(-0.88)		(-0.33)			
200101:200312	-222.49***	0.0342	-212.98***	0.0358	-194.54***	0.0351	980	32,085
	(-7.03)		(-6.72)		(-5.45)			
200201:200412	-164.64***	0.0439	-164.23***	0.0443	-131.00***	0.0469	997	32,640
	(-9.02)		(-9.00)		(-6.71)			
200301:200512	-116.42***	0.0265	-115.72***	0.0265	-124.11***	0.0315	1,034	32,887
	(-7.80)		(-7.75)		(-8.06)			
200401:200612	-97.79***	0.0324	-65.67***	0.0422	-123.41***	0.0477	1,070	33,361
	(-6.54)		(-4.32)		(-8.11)			

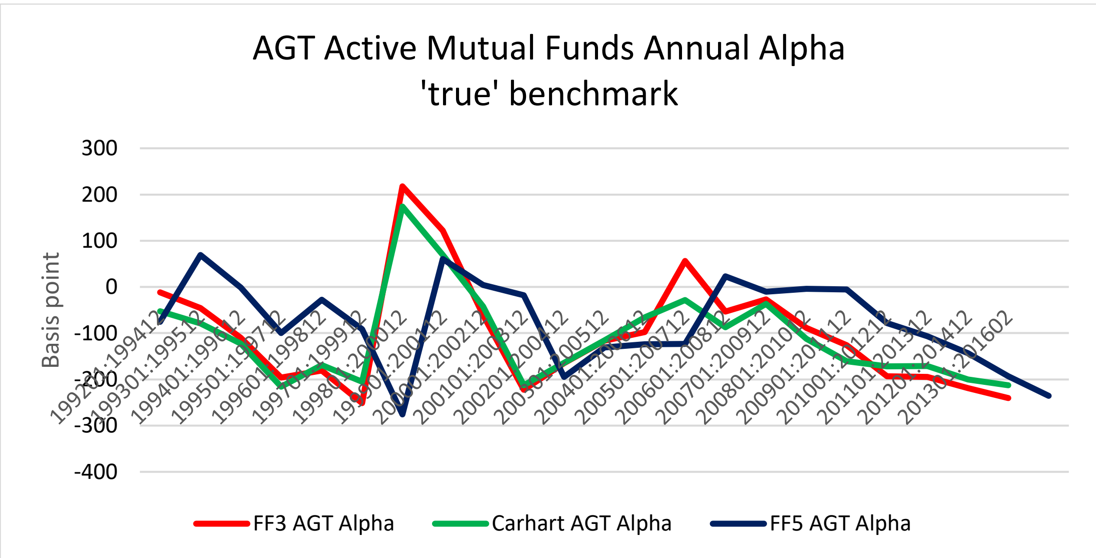
200501:200712	56.10*** (4.43)	0.0161	-28.44** (-2.14)	0.0285	22.79* (1.79)	0.0245	1,066	33,956
200601:200812	-53.60*** (-3.30)	0.0376	-87.02*** (-5.25)	0.0314	-10.66 (-0.59)	0.0151	1,054	34,366
200701:200912	-26.87 (-1.39)	0.0080	-36.42* (-1.85)	0.0099	-4.12 (-0.19)	0.0113	1,057	33,663
200801:201012	-88.51*** (-4.24)	0.0082	-111.97*** (-5.30)	0.0095	-5.59 (-0.25)	0.0118	1,039	32,453
200901:201112	-125.81*** (-6.76)	0.0366	-160.90 (-8.47)	0.0205	-78.02*** (-4.01)	0.0283	975	30,906
201001:201212	-193.09*** (-13.48)	0.0058	-172.09*** (-11.89)	0.0164	-106.33*** (-6.63)	0.0125	895	29,500
201101:201312	-194.56*** (-13.40)	0.0307	-171.19*** (-11.38)	0.0311	-143.52*** (-8.86)	0.0252	855	27,929
201201:201412	-219.28*** (-14.21)	0.0071	-200.04*** (-12.46)	0.0079	-193.34*** (-12.08)	0.0164	789	26,519
201301:201602	-240.61*** (-15.91)	0.0120	-212.50 (-13.66)	0.0144	-235.87*** (-15.50)	0.0161	751	26,573

---

with FF5 versus ‘true’ category benchmark. Overall, alphas calculated for the whole time period are negative and statistically significant for the all three models; however, the performance estimated with FF3 is worse when estimated against S&P500; -74.96 versus -64.92; nevertheless it is better with FF5 and Carhart models, -19.09 and -81.80 versus -75.53 and -83.75, respectively (compared to ‘true’ benchmarks as previously).

The results confirm that the benchmark choice is important and can significantly bias inferences about fund performance. Figure 4 below summarizes the outcomes for AGT alpha (obtained with ‘true’ benchmark) for the three models (Fama and French three and five, and Carhart four factor model) for each of the 22 sub-periods analysed. Overall, all three plots move together, but, as it can be seen from the graph, FF5 provides more extreme values than the other two models and even produces extreme different alphas, as from 1998-2000 to 2000-2002.

Figure 4: AGT alphas (‘true’ benchmark) estimated for the Carhart and Fama and French three and five-factor models.



As the next step we calculate AGT alpha for each mutual fund. Double regressions are performed, as previously, for S&P500 and ‘true’ category benchmarks. Thus, we apply equations (10) and (11) for each fund and each of the 22 rolling sub-periods. Performance is accesses with the use of the Carhart model as the most widely accepted among practitioners. The analysis is conducted for each mutual fund in the sample excluding 460 funds which were assigned to the Large Cap Blend Global category, as their ‘’true’ benchmark is their prospectus benchmark, the S&P500 index. As a result, in total, we estimate 9,393 AGT S&P adjusted alphas, $\alpha_i^{*S\&P500}$, and the same number of AGT 'True' benchmark adjusted alphas, α_i^{*True}.

Table 4: Comparison of average AGT S&P adjusted alphas and average AGT 'true' benchmark adjusted alphas

The table reports comparison of alphas from the following two regressions: $R_{i,t,u} - R_{S\&P500,t,u} = \alpha_i^{*S\&P500} + \beta_{i1,t,u}^{*S\&P500}(R_{M,t} - R_{F,t}) + \beta_{i2}^{*S\&P500} SMB_{t,u} + \beta_{i3}^{*S\&P500} HML_{t,u} + \beta_{i4}^{*S\&P500} WML_{t,u} + e_{i,t,u}^{*S\&P500}$ and $R_{i,t,u} - R_{True,t,u} = \alpha_{i,t,u}^{*True} + \beta_{i1,t,u}^{*True}(R_{M,t} - R_{F,t}) + \beta_{i2}^{*True} SMB_{t,u} + \beta_{i3}^{*True} HML_{t,u} + \beta_{i4}^{*True} WML_{t,u} + e_{i,t,u}^{*}$. $R_{S\&P500,t,u}$ and $R_{True,t,u}$ are the monthly return of the S&P 500 and 'true' Morningstar Global category benchmark respectively (t is the frequency of the data, months and u represents the estimated subperiods in months), $\alpha_i^{*S\&P500}$ is the difference between the Carhart alpha of fund i and its prospectus benchmark, S&P 500, α_i^{*True} is the difference between the Carhart alpha of fund i and the 'true' benchmark; $\beta_{i1}^{*S\&P500}, \beta_{i2,}^{*S\&P500} \beta_{i3,}^{*S\&P500} \beta_{i4}^{*S\&P500}$ are fund i exposures to market risk, size, style and momentum factors beyond the exposure of S&P500 to those risks and $\beta_{i1}^{*True}, \beta_{i2,}^{*True} \beta_{i3,}^{*True} \beta_{i4}^{*True}$ are the fund *i's* four-factor betas adjusted by those of the 'true' benchmark for fund *i*'s category. $R_{M,t,u} - R_{F,t,u}$ is the market risk premium; SMB and HML are size ad value factors from Fama and French (1993) paper and WML is the Carhart (1997) momentum factor, $e_{i,t,u}$ is error term. Alphas and the difference in alphas are annualized and given in bps.

Period	*# of funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} - $\alpha_i^{*S\&P500}$ *(bp)*	*Better* $\alpha_i^{*S\&P500}$ *#/%*	*Better* α_i^{*True} *#/%*
199201:199412	192	43	4	-39	110/58%	81/42%
199301:199512	218	28	-12	-40	121/56%	97/44%
199401:199612	245	-208	-151	57	113/46%	132/54%
199501:199712	275	-356	-217	139	124/45%	151/55%
199601:199812	299	-274	-261	14	143/48%	156/52%
199701:199912	344	-318	-363	-45	165/48%	179/52%
199801:200012	384	341	214	-127	261/68%	123/32%
199901:200112	433	279	198	-81	310/72%	123/28%
200001:200212	476	47	-14	-61	392/82%	84/18%
200101:200312	526	-244	-217	27	311/59%	215/41%
200201:200412	534	-154	-156	-2	287/54%	247/46%
200301:200512	514	29	-111	-140	426/83%	88/17%
200401:200612	524	8	-85	-93	402/77%	122/23%
200501:200712	514	333	78	-255	384/75%	130/25%
200601:200812	513	113	-17	-130	405/79%	108/21%
200701:200912	513	228	2	-226	382/75%	131/25%
200801:201012	506	87	-32	-119	345/68%	161/32%
200901:201112	490	-9	120	129	330/67%	160/33%
201001:201212	487	-225	-178	47	201/41%	286/59%
201101:201312	484	-361	-154	207	55/11%	429/89%
201201:201412	473	-313	-183	130	200/42%	273/58%
201301:201602	449	-330	-231	99	280/62%	169/38%
Overall					5,747/61.2%	3,645/38.80%
*Overall**					2,071/61.18%	1,278/38.16%

*no overlapping period

Table 4 shows the number of unique funds analysed in each period, the average AGT S&P adjusted alphas, $\alpha_i^{*S\&P500}$, and the average AGT 'true' benchmark adjusted alphas, α_i^{*True}, for each of the 22 overlapping periods from January 1992 to February 2016. All alphas are annualised averages across all categories, expressed in basis points. The table also reports the difference between the $\alpha_i^{*S\&P500}$ and the α_i^{*True}. In 15 over 22 periods the $\alpha_i^{*S\&P500}$ are higher than the alternative (column 6 Table 4), implying that using S&P500 as a target instead of a more appropriate benchmark enhances performance. In some periods such as 2000-2002, 2003-2005 and 2006-2008 $\alpha_i^{*S\&P500}$is higher than the α_i^{*True} in at least 79% of funds. In periods such as 2005-2007 and 2007-2009, the average $\alpha_i^{*S\&P500}$ exceeds average α_i^{*True} by over 200 basis points. However, this trend is not as pronounced post financial crises: from 2009-2011 period onwards we find lower percentage of funds (e.g. 11% in 2011-2013) with an average $\alpha_i^{*S\&P500}$ higher than the alternative. In the full (non-overlapping) sample period, deploying the S&P 500 as the AGT adjustment instead of a 'true' category benchmark, overstates the performance for 61.68% of the funds. For the overall sample this difference in AGT adjusted alphas is significant at 1% level using Wilcoxon rank sum test (Z = -5.326).

Although this evidence is pointing that using S&P500 as a benchmark in AGT model results in a better performance for a fund relative to the 'true' benchmark in most of the rolling sub-periods, we do not know whether this benefits more the funds at the top or at the bottom of performance ranks. One should not ignore the fact that there is still 38.8% of the funds in our sample that are worse off by indicating S&P500 as a prospectus benchmark. To further examine the issue of strategic benchmark choices, we investigate whether the fund rankings change considerably when the prospectus benchmark is replaced with a 'true' one.

4.6. The impact of benchmark choice on fund rankings

To conduct the analysis further we examine how the choice of benchmark may impact funds' relative ranking: do winners tend to stay winners and do losers remain losers when the benchmark changes from the one disclosed in the prospectus (S&P 500) to the 'true' benchmark. Using the AGT adjusted monthly alphas for each fund over 22 rolling periods, we split the funds into quartiles according to their performance in each period. Two sets of quartile rankings are constructed one based on AGT S&P500 adjusted alphas and one on AGT 'true' benchmark adjusted alphas. Quartile ranking is not done on a Morningstar global category basis as categories have small number of funds in a number of sub-periods. We construct quartiles using the funds in all the categories excluding those assigned to the Large Cap Blend Global category, as their ''true' benchmark is their prospectus benchmark, the S&P500 index. To examine whether funds move quartiles depending on the benchmark applied we control how many funds remain in the same quartile with S&P500 and 'true' category benchmark. The matching is conducted by fund ID assigned by Morningstar Database. After dividing funds by quartiles we compute average annualised S&P500 and 'true' category benchmark AGT adjusted alphas per each rolling period and each quartile.

Table 5 displays the number of funds in each quartile per rolling period, average annualised AGT alphas (in bps) adjusted for i) prospectus, S&P500 ($\alpha_i^{*S\&P500}$) and ii) 'true' category benchmark (α_i^{*True}) per each rolling period. The table also reports the difference between the two alphas, which signals the magnitude of a possible bias when inappropriate benchmark is used in performance assessment. The last column in Table 5 shows the percentage of funds that remains in the same performance quartile when S&P 500 index is replaced with the 'true' benchmark. AGT-adjusted alphas are estimated with the Carhart model, the results obtained with Fama and French three and five-factor models are qualitatively similar and are presented in the Appendix (Table A1, A2).

Table 5: Difference is alphas per quartile and change of quartile ranks

Panels A-D report results for Quartile 1(top) - 4 (bottom) respectively. All panels show the number of funds and comparison of AGT adjusted alphas, when S&P 500 is used as a benchmark ($\alpha_i^{*S\&P500}$ from eq(5)) and when 'true' benchmark is used (α_i^{*True} from equation (6)). Alphas and the difference in alphas are annualised and given in basis points. The last column shows percentage of funds that remains in the same quartile when the benchmark is changed from the S&P500 to the 'true' benchmark. In the last row, the 'average' represents the average across the periods and across the funds.

Panel A: Quartile 1 (Carhart model)

Period	*# of Funds*	*Average $\alpha_i^{*S\&P500}$ (bp)*	*Average α_i^{*True} (bp)*	*Average difference α_i^{*True} -$\alpha_i^{*S\&P500}$ (bp)*	*% Funds remaining in Quartile 1*
199201:199412	48	724.622	619.908	-104.713	77.08
199301:199512	55	900.993	659.124	-241.869	78.18
199401:199612	61	509.698	550.235	40.537	72.13
199501:199712	69	406.961	518.888	111.927	71.01
199601:199812	75	401.890	450.835	48.945	50.67
199701:199912	86	710.099	637.692	-72.407	75.58
199801:200012	96	1513.924	1414.484	-99.440	68.75
199901:200112	108	1431.359	1279.641	-151.718	76.85
200001:200212	119	1079.855	1310.632	230.777	64.71
200101:200312	132	744.157	987.541	243.384	68.94
200201:200412	134	477.240	521.377	44.137	65.67
200301:200512	129	649.942	438.095	-211.847	72.09
200401:200612	131	643.598	545.615	-97.982	78.63
200501:200712	129	667.274	442.562	-224.712	68.22
200601:200812	128	675.317	402.982	-272.335	75.00
200701:200912	128	1049.981	589.334	-460.646	58.59

200801:201012	127	876.886	556.633	-320.253	77.95
200901:201112	123	586.092	623.681	37.589	61.79
201001:201212	122	390.537	446.654	56.1171	68.03
201101:201312	121	406.372	420.2614	13.889	71.07
201201:201412	118	496.183	386.647	-109.535	69.49
201301:201602	112	234.353	274.290	39.937	71.43
			Average	**-68.19**	**70.09**

Panel B: Quartile 2 (Carhart model)					
Period	***# of Funds***	***Average*** $\alpha_i^{*S\&P500}$ ***(bp)***	***Average*** α_i^{*True} ***(bp)***	***Average difference*** α_i^{*True} -$\alpha_i^{*S\&P500}$ ***(bp)***	***% Funds remaining in Quartile 2***
199201:199412	48	66.766	98.823	32.057	43.75
199301:199512	54	20.647	101.976	81.329	53.70
199401:199612	61	-64.974	-3.744	61.229	52.46
199501:199712	69	-47.359	19.710	67.069	44.93
199601:199812	75	-83.858	-70.943	12.916	9.33
199701:199912	86	-103.845	-93.813	10.033	61.63
199801:200012	96	285.842	277.472	-8.370	54.17
199901:200112	108	284.427	246.554	-37.873	64.81
200001:200212	119	261.234	126.598	-134.636	64.71
200101:200312	131	-93.440	-93.051	0.389	64.89
200201:200412	133	-22.030	-19.082	2.948	45.11
200301:200512	128	117.070	-17.686	-134.756	49.22
200401:200612	131	197.198	66.072	-131.125	54.96
200501:200712	128	242.302	73.913	-168.389	57.03
200601:200812	128	84.538	44.437	-40.100	71.09
200701:200912	128	261.924	86.677	-175.248	32.03
200801:201012	126	47.480	46.723	-0.757	65.87
200901:201112	122	22.702	-65.162	-87.864	63.11
201001:201212	122	-17.213	-13.158	4.055	61.48
201101:201312	121	-37.196	13.985	51.181	71.07
201201:201412	118	-20.476	-72.422	-51.946	67.80
201301:201602	112	-121.013	-99.772	21.241	67.86
			Average	**-28.48**	**55.50**

Panel C: Quartile 3 (Carhart model)					
Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} -$\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 3*
199201:199412	47	-187.715	-147.318	40.397	38.30
199301:199512	54	-224.299	-133.920	90.379	51.85
199401:199612	62	-306.194	-294.377	11.817	50.00
199501:199712	68	-352.285	-286.392	65.893	52.94
199601:199812	74	-357.022	-350.239	6.783	25.68
199701:199912	86	-402.512	-377.958	24.555	66.28
199801:200012	96	-98.842	-100.839	-1.997	55.21
199901:200112	109	-85.088	-109.625	-24.537	75.23
200001:200212	119	-119.334	-260.565	-141.231	70.59
200101:200312	131	-446.355	-465.370	-19.015	64.12
200201:200412	133	-261.305	-285.984	-24.679	51.88
200301:200512	128	-105.933	-239.287	-133.353	49.22
200401:200612	131	-20.324	-169.740	-149.416	54.96
200501:200712	128	7.973	-116.063	-124.036	51.56
200601:200812	129	-129.902	-165.936	-36.034	78.29
200701:200912	129	19.864	-129.194	-149.058	34.11
200801:201012	126	-157.862	-179.444	-21.581	66.67
200901:201112	122	-259.939	-334.760	-74.821	55.74
201001:201212	121	-251.755	-234.360	17.395	67.77
201101:201312	121	-279.317	-205.105	74.211	66.94
201201:201412	119	-241.206	-273.286	-32.080	69.75
201301:201602	113	-320.076	-282.608	37.468	64.60
			Average	**-25.59**	**57.35**

Panel D: Quartile 4(Carhart model)					
Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} -$\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 4*
199201:199412	48	-680.588	-695.340	-14.752	72.92
199301:199512	55	-780.702	-735.892	44.809	69.09
199401:199612	61	-930.347	-884.578	45.769	67.21
199501:199712	69	-1436.729	-1173.304	263.425	69.57
199601:199812	75	-1073.939	-941.362	132.577	64.00
199701:199912	86	-1218.210	-1248.937	-30.727	80.23
199801:200012	96	-743.990	-740.403	3.588	68.75
199901:200112	108	-709.590	-800.757	-91.167	76.85
200001:200212	119	-839.593	-890.372	-50.779	68.91
200101:200312	132	-1038.163	-1096.031	-57.868	77.27
200201:200412	134	-770.944	-797.378	-26.434	79.10
200301:200512	129	-492.279	-635.970	-143.691	68.22
200401:200612	131	-445.152	-626.243	-181.091	78.63
200501:200712	129	-352.865	-477.177	-124.312	68.99
200601:200812	128	-517.257	-523.115	-5.858	80.47
200701:200912	128	-495.099	-480.083	15.015	53.91
200801:201012	127	-651.181	-620.318	30.863	81.10
200901:201112	123	-778.481	-777.586	0.895	65.04

201001:201212	122	-830.054	-835.978	-5.924	80.33
201101:201312	121	-1193.792	-802.262	391.529	69.42
201201:201412	118	-1137.554	-828.766	308.788	72.88
201301:201602	112	-832.393	-611.258	221.135	74.11
			Average	**32.99**	**72.14**

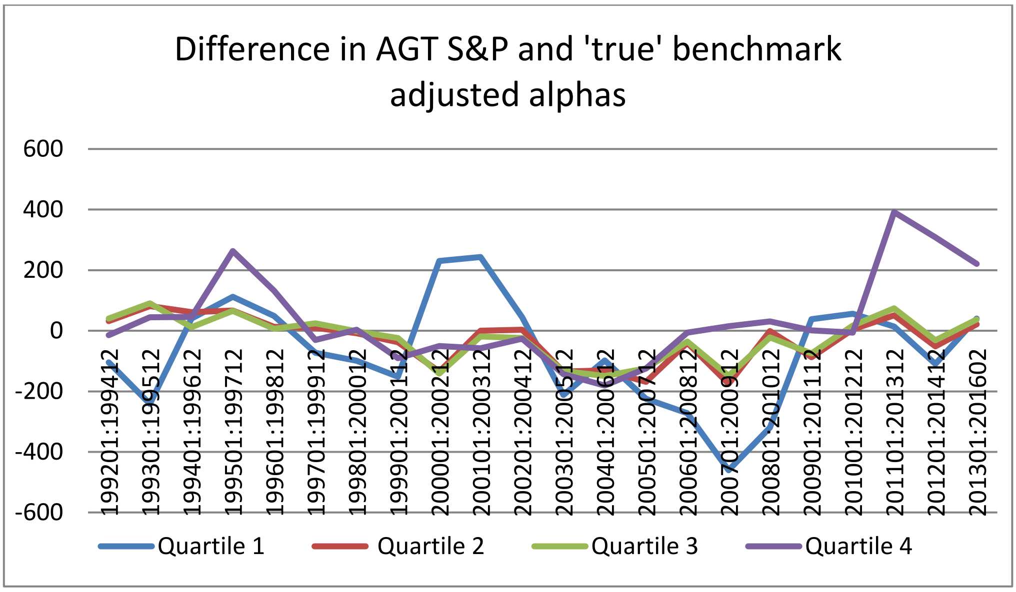
In Panel A (Quartile 1) 'true' benchmarks provide a lower average AGT adjusted alpha in 12 out of 22 periods analysed, indicating that for 55% of the periods performance of winners estimated with S&P500 is overstated. Analogous tendency can be viewed for the Quartile 2, 3 and 4 (11, 13 and 11 out of 22 periods of lower average AGT 'true' benchmark adjusted alphas, α_i^{*True}, respectively).

More importantly, the average number of funds that remains in the top quarter over the years (70%), implies that 30% of the top performing funds drop in performance ranks and leave the quartile when the performance is adjusted with the 'true' benchmark. Overall, for the total sample period, there is on average 68bps advantage for Quartile 1 funds of using S&P 500 as the prospectus benchmark. Comparing this value to the equivalent average alpha difference in Panels B-D, it becomes evident that the top performing funds benefit most from the choice of prospectus benchmark. Panel D in fact suggests that Quartile 4 funds get penalised from inadequate benchmark selection. Thus, on average, close to 30% of funds move up in quartile rankings when their performance is assessed against a 'true' global category benchmark. The average 'true' AGT adjusted alpha for the total period is 33 basis points higher than the one estimated with prospectus benchmark, leading us to conclude that these funds would be better off selecting a 'true' benchmark.

Quartile 2 and Quartile 3 funds (Panel B and C of Table 5) are of least interest to investors; the funds in these quartiles are neither the top funds investors look out for nor the ones at the bottom they are trying to avoid. However, we document that the results for both quartiles are similar: adjusting alphas with the 'true' benchmark changes, on average, the quartile ranking of 45% and 43% of funds from Quartile 2 and Quartile 3, respectively. Those movements can be in both directions – up to a higher or down to a lower ranked quartile, and in most of the cases there is an interchange between these two groups. The AGT alpha adjusted 'true' with the true benchmark is on average 28 (Quartile 2) and 26 (Quartile 3) basis points higher than the one adjusted with self-declared prospectus benchmark.

Therefore, inferences on mutual fund relative performance may be significantly biased when fund performance is evaluated in respect to unsuitable benchmark. To support our discussion, we plot the difference in average AGT adjusted alphas, $\alpha_i^{*S\&P500}$ α_i^{*True} in each ranking quartile and each period (column 5 from Table 5) in Figure 5.

Figure 5: Difference between AGT S&P adjusted and AGT 'true' benchmark adjusted alphas



The figure shows that the average AGT adjusted alphas for the Quartile 2 and 3 are almost identical irrespective of the benchmark. However, the performance of top funds is overestimated with the prospectus benchmark in over half of the rolling periods. The difference in AGT adjusted alphas for Quartile 1 funds reaches peaks of -241bps in 1993-1995, around -300 bps in 2006-2008 and 2008-2010, and a maximum of -460bps as in 2007-2009, in favour of alphas adjusted with the S&P rather than the 'true' benchmark. Even though top performing funds seem to take advantage of using S&P500 as their benchmark; there are also cases when performance of these funds benchmarking against prospectus benchmark could be undervalued, as in 2000-2002 and 2001-2003, but by a smaller margin. In contrast, our results show, that while benchmarking against prospectus benchmark is on the whole beneficial to winners, it negatively affects the performance of losers. The difference in AGT adjusted alphas of Quartile 4 funds in some time periods, for instance 2011-2013 and 2012-2014, reaches 392 and 309 basis points, respectively, in favour of alphas adjusted with the 'true' benchmark.

Considering these findings, most of the funds that are potentially strategically selecting S&P 500 as the benchmark and benefiting from it are those in top performance quartile. They have on average 0.68% higher benchmark-adjusted alphas when that benchmark is the one given in the prospectus and nearly 30% of those funds lose the 'winner' status when the self-declared benchmark is substituted with a better suited one. In all other quartiles there is no clear advantage of using S&P 500 as a prospectus benchmark. Hence, the choice of the benchmark affects not only the inferences about a fund's absolute performance, but it can also mislead investors about its relative performance. This leads us to conclude that any information in fund prospectus about the performance relative to the prospectus benchmark or relative to other funds should be treated with caution.

4.7. Conclusions

This paper analyses the impact of benchmark choice on US equity funds performance and gauges potential biases in absolute and relative performance evaluation stemming from the inaccurate prospectus benchmark selection. We re-visit the question of mismatch between the prospectus benchmark and fund objectives, raised in Sensoy (2009), and estimate the impact of such misclassification on fund performance and ranking. In contrast to the previous literature, our analysis accounts for non-zero benchmark alphas produced by standard pricing models, discussed in recent literature such as Chan, Dimmock, and Lakonishok (2009), Cremers, Petajisto and Zitzewitz (2012). Our sample includes net monthly returns of 1281 actively managed US equity mutual funds from January 1992 to February 2016, that report the S&P 500 index as their primary prospectus benchmark in the Morningstar database. We find that only 460 of those funds belong to the Large Cap blend Morningstar category, for which the S&P 500 would be the most suited benchmark. All other remaining funds fall across 21 other distinct Morningstar Global categories, some of which imply that fund risk profile and composition is very different from that of their prospectus benchmark. Naturally, we investigate whether the fund's performance relative to the S&P 500 is better than when measured against what we consider their 'true' category benchmark. Regression of mutual fund returns on the returns of S&P500 and the 'true' Morningstar Global category benchmark a fund belongs to, shows that 'true' category benchmarks are a better fit for our funds, having on average 10% higher R-squared.

Further, in our preliminary analysis, similar to Cremers et al. (2012) and Chinthalapati et al. (2017), we report non-zero alphas of passive benchmark indices in our sample. To eliminate the upward/downward biases in performance assessment caused by embedded benchmark alphas we apply Angelidis et al. (2013) method (AGT) that adjusts fund's alpha for benchmark's alpha, hence isolating manager's skill above that common to the benchmark. Performance for each fund is calculated against the prospectus and the 'true' benchmark, more appropriate for the Morningstar category a mutual fund belongs to. The sample period is split into 22 rolling overlapping windows, each being 36 month in length. Overall, for the total sample period we document higher AGT four-factors alphas estimated with S&P 500 as a benchmark versus those adjusted with the 'true' category benchmark. Fund performance estimated with the Carhart model for each mutual fund with 36 month period revealed that in 70 percent of the periods the average AGT alphas adjusted with S&P500 are higher and overestimate fund performance. Overall figures for the entire period show that 61.2% of the funds benefit from wrongly benchmarking their performance against S&P500 (prospectus benchmark). So that, the average AGT-adjusted alpha drops by 23 basis points when 'true' global categories benchmark is used.

Additional results show that inaccurate benchmark choice also influence relative performance assessment. Thus, 30% of top performing funds move their ranking position when their performance is adjusted with the 'true' benchmark. Surprisingly, the results show that the worst performing funds get penalised by stating inaccurate benchmark in their prospectus. In fact, close to 30% of losers move up their position when performance is estimated with the most suitable global category benchmark. The results indicate that the top quartile funds benefit most from the choice of prospectus benchmark. Thus, for instance, in 2007-2009 the difference in S&P 500 and 'true' benchmark adjusted alphas reached 460 bps in favour of using the prospectus benchmark. This leads us to conclude that strategic benchmark selection appears to be most likely in the funds at the top performance quartile, while we do not observe clear advantage of benchmark gaming in the remaining quartiles. Overall, we can highlight that the average alpha when the performance is adjusted with 'true' Global category benchmarks drops by 68bps in Quartile 1, falls 28bps and 25bps in Quartiles 2 and 3, and increases 33bps in Quartile 4 in the whole sample period.

These results show that appropriate benchmarking is essential for accurate performance evaluation, as inferences on both fund performance and performance ranking may change significantly when estimated against a 'true' benchmark instead of their self-declared prospectus benchmark. It is irrefutable in this paper that information disclosed by equity mutual funds regarding fund's self-designated benchmark is by and large not accurate as prospectus benchmarks do not represent funds' actual investment style in 2/3 of the funds in our sample. This study raises concerns that require attention of financial regulators and policy makers. New information disclosure requirements should be placed to provide more clarity for investors as to how the prospectus benchmark is selected. It also calls for investors to be more cautious when interpreting performance figures in fund prospectus. The paper can be extended to non-equity funds or funds where the benchmarking is ambiguous (such as hedge funds for instance).

Chapter 5

Third empirical essay

Abstract

We assess UK mutual fund performance from a perspective of a peer-group, applying a novel approach suggested in Hunter et al. (2014). Our sample comprises of 817 UK long-only active equity mutual funds allocated to nine Morningstar style category peer-groups in the period 1992-2016. Overall, we find that those funds with most significant positive peer-group adjusted alphas continue to perform well one-year-ahead, using both parametric and non-parametric measures of persistence in performance. Further, a small increase in significance of peer-group adjusted alphas significantly improves the probability that a fund will be placed in the top quartile in the following period. Finally, we document that persistence in performance is driven by both winner and loser funds. The results within each peer group by and large conform to these findings.

Keywords: UK Equity Mutual Funds, Active Peer Benchmark, Performance ranking, Performance persistence

JEL classification: G11, G12, G23

5.1. Introduction

Recent Financial Conduct Authority (FCA) study on the UK asset management market[28] reveals some findings that can be a cause for concern of both UK retail and institutional investors. They find that neither active nor passive funds manage to outperform their self-reported benchmarks after fees. Further, they report lack of clarity on how funds select their prospectus benchmarks and concern that such (poorly) selected benchmarks can lead to misinterpretation of fund performance. The issue of inappropriately chosen funds' self-reported benchmarks is documented in Sensoy (2009) and Kacperczyk, Sialm, and Zheng (2008) for the US market. In this UK study, we note a similar issue. Our sample comprises of 817 UK active equity long-only mutual funds that Morningstar allocates to nine different style categories, yet nearly 65% of those report FTSE All Share Index as their benchmark. However, the problem arises for a UK investor interested in, say, small-cap growth stocks – the off-the-shelf style specific index by a standard index provider (FTSE, MSCI) for this group does not exist. The same is true for other value and growth style categories.

On the other hand, the academics have adopted the Capital Asset Pricing Model, the Fama and French three- and five-factor models and the Carhart four-factor model as standard benchmarks for performance evaluation. However, each of these different approaches can lead to very different inferences about fund performance. What is more, they tend to produce non-zero alphas for passive benchmark indices used by many funds as their prospectus benchmarks. For example, Cremers, Petajisto and Zitzewitz (2012) find annual Carhart alpha of the S&P 500 index of 0.82 % (significant at 1%) from 1980 to 2005. Significant non-zero alphas are presented in Costa and Jakob (2006).

From this, it is evident that even if a UK fund identifies the benchmark appropriate for their objectives, the benchmark index itself may generate positive alphas in the standard factor models. Hence, a fund manager that simply replicates such benchmark will be classified as skilful, whereas in fact exhibits a median performance. Therefore, the first question this discussion raises is what would investors accept as a suitable benchmark in the UK fund performance measurement, given that style specific benchmarks are not available? In this paper, we take the view that investors are driven by their aptitude towards risk/reward and are

[28] FCA Asset Management Market Study, Final Report (June, 2017): https://www.fca.org.uk/publication/market-studies/ms15-2-3.pdf

unlikely to consider investments across investment styles in different risk categories: i.e. a typical value investor will not be interested in the growth fund or its performance, but would be interested in identifying the value fund in the category that would give them highest rewards. The acceptable benchmark should reflect this view. Second, if the acceptable benchmark can be identified, how should it be accounted for in the performance measurement models? An third, if the out/underperformance is identified amongst the UK funds, can it persist? We address these questions and contribute to the literature of UK equity mutual fund performance and persistence, from the perspective of a peer group; a group of funds with similar investment style and objectives. We use approach by Hunter, Kandel, Kandel and Wermers (2014) that enables us to account for commonalities across funds strategies within a peer group and identify funds with best risk-adjusted performance within a group. In essence, this approach provides investors with the peer-group adjusted alphas. The method is novel in the finance literature; it has not been tested in the UK market and has wide practical implications, of significance to both individual and institutional investors. It reveals the best performing funds within the reference group of equal style and risk exposure.

The literature on UK fund performance measurement is scarce compared to the US. A number of studies agrees that there is no strong evidence of outperformance, see for instance Cuthbertson, Nitzsche and O'Sullivan (2008, 2012) for more recent evidence or Blake and Timmermann (1998) for earlier evidence. These studies apply standard multi-factor models to measure performance and do not adjust for presence of non-zero alphas in passive benchmarks. Angelidis et al. (2012) approach enables investors to compute the benchmark-adjusted alpha for a fund (AGT alpha hereafter), defined as alpha of the fund minus the alpha of the self-reported benchmark. Therefore, AGT alpha will be useful to assess relative rankings of funds within a peer group only if 1) all the funds in the group have the same self-reported benchmark and 2) if that benchmark is appropriately selected to reflect funds' characteristics. In our sample, neither 1) nor 2) holds within Morningstar style peer groups. This warrants Hunter et al. (2014) to be more suitable methodology for our paper. Further, most of the existing studies on UK mutual fund performance do not differentiate between investment styles. One recent study for instance, Otten and Reijnders (2012), finds that UK small cap funds generate statistically significant alpha of 4.08% p.a., net of fees. This is also at odds with previous studies in developed markets, which have predominantly a large cap focus. Therefore, the performance across different style groups is worth exploring further.

Studies on persistence in fund performance in the UK are even fewer and provide mixed evidence. For instance, Blake and Timmerman (1998) find that there is some, but not overwhelming evidence of persistence, while Quigley and Sinquefield (2000) find that only poor performance persists. Fletcher and Forbes (2002) investigate persistence in UK mutual funds by applying different models: CAPM, APT and the four-factor Carhart (1997) model. The work verifies significant performance persistence for portfolios formed on the basis of prior year excess return, when persistence is evaluated with CAPM and APT; but it disappears when using the four-factor model. However, conditional performance measure of Ferson and Schadt (1996) reversed this result with even stronger evidence of statistical significance. This shows that different benchmark models lead to different conclusions about persistence. The study of Hunter et al. (2014) shows that their peer-group adjusted alpha model has better ability to select future winner funds than the standard Carhart model, in the US. In this paper, we investigate the predictive ability of the peer-group adjusted alpha model in the UK market. Specifically, we test whether selecting the funds with highest peer-group-adjusted performance will enable investors to earn superior excess returns and four-factor alphas one-year-ahead.

The methodology is applied to a group of 817 active UK equity long-only mutual funds in the period January 1992 to February 2016. The funds are split into three Morningstar categories by size (Small, Medium Large) and further three categories by style (Value, Blend, Growth); nine categories in total. Hunter et al. (2014) propose an Active Peer-group Benchmark (APB) as a benchmark for performance evaluation, which is an equally weighted portfolio of funds within the same peer group. We consider the nine Morningstar categories as peer groups. The Active Peer Benchmark's Carhart alpha and Carhart error term are estimated and included as additional factors to the standard Carhart model when evaluating performance of a fund. If a fund manager has skill superior relative to common idiosyncratic risk taken, the APB-adjusted alpha in the new, APB adjusted model will be positive and significant.

We find that ABP adjusted model has higher R-squared when fitted to fund returns than the standard Carhart model. The alphas of the two models are found to be different: in 55% of the cases APB adjusted alphas are higher, but that bares no real implications for investors. What is of importance is whether these adjusted alphas obtained from a better-fitted model are indeed enabling us to select future winners. To test for this persistence, we split the funds into four performance quartiles. We apply several tests of performance persistence, parametric (regression) and non-parametric (contingency tables). All tests agree that APB adjusted alpha is a strong predictor of performance one-year-ahead in the UK. In our full sample period, 1% increase in of monthly APB alpha obtained using 36 months of historical data leads to 15 basis points rise in alpha one year forward, a result significant at 1% level. Also, our ordered probity model shows that a 1% increase in peer-group adjusted alpha's t-statistics increases the probability for a fund to be placed in the top performance quartile by 2.4% (significant at 1%).

The paper is organised as follows: Section 2 presents the data and methodology. Section 3 shows preliminary results. Section 4 reveals the predictive ability of ABP adjusted alphas and Section 5 concludes the paper.

5.2. Data and Methodology

5.2.1. Data Description

The sample comprises of 817 active UK long-only equity mutual funds. The sample period spans from January 1992 to February 2016 and includes 125,305 net monthly total return observations for the funds (inclusive of dividends). We split the funds according to the category Morningstar assigns them to. There are three size categories in Morningstar (Large, Medium and Small Cap) and within each size category there are three style categories (value, blend and growth), a total of nine (3x3) categories overall. Table 1 shows the number of funds and monthly observations, together with percentage of funds and percentage of monthly observations in each category. Fund net returns and their Morningstar categories are obtained from Morningstar.

Table 1: Sample and Peer-Group Categories (Morningstar style box)

The sample consists of 817 (125,305 monthly observations) long-only active UK equity mutual funds from January 1992 to February 2016. Table below shows the number of funds and number of monthly observations per style investment provided by Morningstar.

Morningstar Style	# Funds	# Monthly Observations	Percentage Funds	Percentage Monthly Observations
Large Value	222	35,007	27.17	27.94
Large Blend	310	46,805	37.94	37.35
Large Growth	48	7,604	5.88	6.07
Mid Value	28	4,949	3.43	3.95
Mid Blend	68	8,755	8.32	6.99
Mid Growth	28	3,957	3.43	3.16
Small Value	12	1,886	1.47	1.51
Small Blend	42	5,790	5.14	4.62
Small Growth	59	10,552	7.22	8.42

Total	817	125,305		

Over 60% of funds are concentrated in the Large Value and Large Blend category. Only 1.47% are placed in the Small Value style, that according to the literature are the styles that have historically outperformed their counterparts[29], at least in the long run.

Our sample period is split into 21 rolling (overlapping) 36-monthly periods. The first rolling window is Jan 1992-Dec 1994, the second is rolled forward by one year from Jan 1993-Dec 1995 and so on, until the end of the sample period is reached[30]. For performance estimation and construction of the active peer benchmark, we require minimum of 36 months of continuous observations. This implies that even if a fund has at least 36 months of returns in total, in each rolling window the number of observations may be less than that; however, we restrict the number of observations in each rolling window to be minimum 30 months. The number of funds that meets the criteria falls to 780.

5.2.2. Active Peer-Group Benchmarking methodology

We follow the new performance evaluation methodology proposed by Hunter et al. (2014). The approach modifies the standard factor models, such as Fama-French three- and five-factors and Carhart four factor models, by adding new information related to the benchmark. The authors refer to it as an "Active Peer Benchmark" (APB), which can be viewed as a passive benchmark for a fund. Passive benchmark indices commonly used as performance targets for active funds are associated with two main issues in the recent literature. First, following, for instance, Cremers et al. (2012) and Chinthalapati et al. (2017), well known passive benchmarks have non-zero alphas. Second, following Sensoy (2009), many funds do not choose as their prospectus benchmark the passive index that best fits their investment strategy. In our sample, nearly 65% of the funds report FTSE All Share index as their benchmark, including 17.5% of funds from the small cap Morningstar style category. Hence, the problem of benchmark-style mismatch is present in our sample too.

[29] See for instance Chan and Lakonishok (2004), Dimpson, Marsh and Staunton (2004), Fama and French (1998), Reinganum (1999) among many others for evidence on small cap and value outperformance.
[30] Note that the last rolling period has 38 months as the sample ends in February 2016. Also note that in Tables 4, 5, 6 the last rolling period used for prediction of future performance is 2012-2014.

Using APB as a passive benchmark for a fund, we overcome both of these issues. First, the peer-group in our study is defined as the Morningstar style category that a fund belongs to, and by definition, the funds are placed in a category by Morningstar according to their holdings and risk profile. Second, in Hunter et al. (2014) model, it is not relevant if the alpha of APB is positive, negative or zero, what matters is whether the fund has done better/worse than this benchmark. Hence, the main intuition in this approach lies in adding the APB to the standard Carhart four factors in this study to enable investors to account for the benchmark in the model and estimate funds' alphas relative to their peer-group. Let us lay out the steps of this approach.

We start by choosing a baseline model for fund performance measurement. To this end, we opt for the standard Carhart (1997) four-factor model, commonly used in the literature on UK fund performance:

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{M,i,t}\left(R_{M,t} - R_{F,t}\right) + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,WML}WML_t + e_{i,t} \quad (1)$$

Where $R_{i,t}$ is the return of a mutual fund i in period t, R_F is the UK risk free rate; $R_{M,t} - R_{F,t}$ is the UK market risk premium; SMB and HML are Fama-French size and value factors and WML is the Carhart (1997) momentum factor. The market risk premium[31], the risk free rate, SMB, HML and WML factors are all defined as per Gregory, Tharyan and Christides (2013) paper and obtained from University of Exeter, Xfi Centre for Finance and Investment website[32]. α_i represents the four-factor alpha of fund i, i.e. the excess return of the fund unexplained by the four risk factors, $e_{i,t}$ is error term.

Next, we estimate the APB excess return, where $R_{APB,i,t}$ is the return of active peer group to which fund i is allocated to; $R_{F,t}$ is risk free rate. The APB excess return is defined as the equally weighted average excess return of all the mutual funds in the same Morningstar peer-group category:

$$R_{APB,i,t} - R_{F,t} = \frac{1}{N_{APB,i}} \sum_{i=1}^{N_{R_{APB,i,t}-R_{F,t}}} R_{i,t} \quad (2)$$

[31] UK market risk premium represents the return on FTSE All Share Index (R_M) minus one month UK Treasury bill (R_F), as per Gregory et al. (2013)
[32] http://business-school.exeter.ac.uk/research/areas/centres/xfi/research/famafrench/files/

Where $N_{APB,i}$ is the number of funds in the given Morningstar peer-group. We use nine peer-group categories as per Table 1, and, consequently, construct nine APBs.

In Hunter et al. (2014) APB-adjusted alpha model, the APB is used as the augmentation factor for the standard Cahart (1997) model to account for commonalities of mutual fund strategies in the same peer group and isolate the unique fund manager's skill. Hence, to begin with, the APB excess return ($R_{APB,i,t} - R_{F,t}$) is regressed against the standard Carhart four-factors:

$$R_{APB,t} - R_{F,t} = \alpha_{APB,t} + \beta_{APB,M,t}\left(R_{M,t} - R_{F,t}\right) + \beta_{APB,SMB}SMB_t + \beta_{APB,HML}HML_t + \beta_{APB,WML}WML_t + e_{APB,t} \quad (3)$$

where α_{APB} is the alpha of the Active Peer-group Benchmark and $e_{APB,t}$ is the APB residual. All else is as described earlier in this section.

As the next step, the estimated $\widehat{a_{APB,t}}$ and $\widehat{e_{APB,t}}$ values are included to the Carhart four factor model to adjust for the common alphas within a peer-group:

$$R_{i,t-R_{F,t}} = \alpha_{i,ADJ} + \beta_{i,M,t}\left(R_{M,t} - R_{F,t}\right) + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,WML}WML_t + \beta_{i,ADJ}\left(\widehat{a_{APB,t}} + \widehat{e_{APB,t}}\right) + \omega_{i,ADJ,t} \quad (4)$$

Here, $\widehat{a_{APB,t}} + \widehat{e_{APB,t}}$ is the adjustment factor, the new $\alpha_{i,ADJ}$ is the APB-adjusted alpha, which reflects unique fund manager's skill and takes away any performance that may be the result of a manager undertaking an investment approach and risks common for the peer group.

As has been mentioned above in this chapter we follow the methodology of Hunter et al. (2014). Thus, first, we estimate the excess return of each peer group for every 36 months rolling periods (21 periods) following equation (2). Second, we estimate $\alpha_{APB,t}$ (alpha of the peer group) for every 36-month time period by regressing the excess return of each peer group ($R_{APB,t} - R_{F,t}$), 9 styles, versus the Carhart four factor model. Third, we use obtained peer group alphas $a_{APB,t}$ and estimated betas $\beta_{i,M}, \beta_{i,SMB}, \beta_{i,HML}, \beta_{i,WML}$ from each regression and calculate the return of each peer group predicted by the model (R_{APB}expected), for each period. Fourth, we calculate the difference between the real and expected APB returns and obtain APB residuals for every peer group, 9 styles, and all 21 periods ($e_{APB,t}$). Fifth, we combine estimated $\alpha_{APB,t}$ (alpha of the peer group) and $e_{APB,t}$ (APB residuals) and include as the fifth factor to the Carhart model. Hence, we obtain the Hunter et al. (2014) APB-adjusted alpha model for each peer group (9 styles) and each period.

By regressing the excess return of each mutual fund ($R_{i,t}$-$R_{F,t}$) versus style-corresponding APB-adjusted model (regression 4) we find the new $\alpha_{i,ADJ}$ (APB-adjusted alpha) which represents a unique fund manager's skill which is above/uncorrelated with the managers' peer group average skills and cannot be explained by the peer group commonalities in idiosyncratic risk taking. The results have been estimated for each rolling period.

5.3. Preliminary Results:

5.3.1 Active Peer Benchmark Alphas

Table 2 shows Carhart four-factor (monthly) alphas for each of the nine peer-group benchmarks, obtained using equation (3). They are calculated for each of the 21 rolling overlapping sub periods and reported here together with the t-statistics and R-squared of the model. The last row reports the alphas for the full sample period for each category.

Many of the reported alphas are negative, but they are also overwhelmingly non-significant across periods and peer-group categories. The instances in which alphas are significant are linked to well-documented periods of out/underperformance of certain groups. For instance, Small-Cap (Growth and Blend) and Mid-Cap Growth category generate significant positive alphas during the dot.com boom period. Further, most of the significant alphas are found in the Small-Cap categories (which goes in line with literature on small-cap outperformance) and a few in the Large-Cap Value and Mid-Cap Growth category.

These, overall, by and large insignificant alphas are expected: within each peer group, some funds have positive while some have negative alphas which balance out by construction of the APB. Given this, our APBs fit the standard passive benchmark definition, such as that given by Chen and Knez (1996), who state that if a benchmark is used in performance measurement it should generate no excess performance itself. The aim of this paper is, therefore, to utilize such benchmark to identify the best and persistent outperformers, within each peer group.

Table 2: Active Peer-group Benchmark (APB) alpha

The table shows the Carhart alpha of the Active peer-group benchmark (APB), its t-statistics and R-squared, all obtained from the following model: $R_{APB,t,u} = \alpha_{APB} + \beta_{APB,M,t,u}(R_{M,t,u} - R_{F,t,u}) + \beta_{APB,SMB}SMB_{t,u} + \beta_{APB,HML}HML_{t,u} + \beta_{APB,WML}WML_{t,u} + e_{APB,t,u}$. In the model, $R_{APB,t,u}$ is the monthly excess return of the peer group in period *u* is defined as the equally weighted average excess return of all mutual funds in the same peer-group category (*t* is the frequency of the data, months and *u* represents the estimated subperiods in months); $R_{M,t,u} - R_{F,t,u}$ is the UK market risk premium; SMB and HML are Fama-French size and value factors and WML is the Carhart (1997) momentum factor. The market risk premium, the risk free rate, SMB, HML and WML factors are all defined as per Gregory, Tharyan and Christides (2013); α_{APB} is the alpha of the APB and $e_{APB,t}$ is the APB residual. Peer-group is defined as the Morningstar style category. Numbers in **bold** mark the results for the total sample period.

Period		Large Value	Large Blend	Large Growth	Mid Value	Mid Blend	Mid Growth	Small Value	Small Blend	Small Growth
199201:199412	α_{APB}	-0.00066	-0.0003054	-0.000543	0.0017076	-0.0006478	0.0004054	-0.0005753	0.0015277	0.0004322
	(t-stat)	(-0.39)	(-0.30)	(-0.41)	(1.00)	(-0.51)	(0.20)	(-0.38)	(0.58)	(0.20)
	R^2	0.9807	0.9876	0.9816	0.9736	0.9787	0.9716	0.9772	0.9243	0.9590
199301:199512	α_{APB}	-0.0004615	-0.0007108	-0.000611	0.0009425	-0.0004764	-0.000617	0.0011894	0.0018284	0.0030435
	(t-stat)	(-0.38)	(-0.75)	(-0.52)	(0.61)	(-0.38)	(-0.30)	(1.06)	(0.94)	(1.58)
	R^2	0.9730	0.9825	0.9753	0.9564	0.9625	0.9414	0.9765	0.9321	0.9406
199401:199612	α_{APB}	-0.0004771	-0.0002684	-0.0004102	0.0005906	-0.0002725	0.0002938	-0.0005024	0.0002568	0.0014368
	(t-stat)	(-0.52)	(-0.32)	(-0.37)	(0.52)	(-0.25)	(0.16)	(-0.54)	(0.15)	(0.82)
	R^2	0.9800	0.9835	0.9711	0.9689	0.9640	0.9356	0.9788	0.9293	0.9354
199501:199712	α_{APB}	0.0016871*	0.0005945	0.0015605	0.0019782	0.0000803	0.00000	0.0012172	0.0034281*	0.0008584
	(t-stat)	(1.84)	(0.62)	(1.23)	(1.25)	(0.07)	(0.00)	(0.83)	(1.76)	(0.44)
	R^2	0.9735	0.9721	0.9471	0.9200	0.9382	0.8816	0.9262	0.8892	0.9001
199601:199812	α_{APB}	0.0006589	0.0005399	0.0016424	0.0008898	-0.0003595	0.0009752	-0.0020622	0.0008892	-0.0010623
	(t-stat)	(0.53)	(0.44)	(1.19)	(0.56)	(-0.25)	(0.56)	(-1.13)	(0.39)	(-0.52)
	R^2	0.9721	0.9732	0.9663	0.9567	0.9476	0.9464	0.9475	0.9344	0.9449
199701:199912	α_{APB}	-0.0008416	-0.000921	-0.0003234	-0.0008087	-0.0023043	0.0013975	0.0002996	0.0004128	-0.0001833
	(t-stat)	(-0.64)	(-0.70)	(-0.23)	(-0.43)	(-1.46)	(0.62)	(0.08)	(0.13)	(-0.07)
	R^2	0.9707	0.9722	0.9662	0.9482	0.9412	0.9378	0.8893	0.9148	0.9324
199801:200012	α_{APB}	0.0002936	0.0005477	0.0015383	0.0022651	0.0001254	0.0044112*	0.0059391	0.0088466**	0.0056658*
	(t-stat)	(0.24)	(0.43)	(1.06)	(1.15)	(0.09)	(1.75)	(1.47)	(2.19)	(1.69)
	R^2	0.9727	0.9714	0.9612	0.9349	0.9521	0.9372	0.8847	0.8697	0.9139
199901:200112	α_{APB}	-0.0005645	0.0002673	0.0010598	0.0018821	-0.0004256	0.0047429*	0.0078549*	0.0074173	0.0050906
	(t-stat)	(-0.36)	(0.21)	(0.64)	(0.75)	(-0.23)	(1.81)	(1.90)	(1.64)	(1.32)
	R^2	0.9658	0.9776	0.9630	0.9234	0.9533	0.9565	0.9248	0.8951	0.9261
200001:200212	α_{APB}	-0.000119	0.0003947	0.0011177	0.0020279	0.0009779	0.0040947	0.0014607	0.0044475	0.0016455
	(t-stat)	(-0.07)	(0.31)	(0.60)	(0.73)	(0.61)	(1.61)	(0.49)	(1.43)	(0.53)
	R^2	0.9706	0.9841	0.9681	0.9325	0.9775	0.9582	0.9566	0.9511	0.9541
200101:200312	α_{APB}	-0.0005553	-0.0001454	0.0001513	-0.0002945	-0.0002344	-0.0001922	-0.0036389	0.0006147	0.0000696
	(t-stat)	(-0.35)	(-0.09)	(0.08)	(-0.15)	(-0.14)	(-0.09)	(-1.43)	(0.22)	(0.03)
	R^2	0.9810	0.9809	0.9747	0.9739	0.9834	0.9711	0.9705	0.9642	0.9664

200201:200412	α_{APB}	0.0001256	-0.0006789	0.0011444	0.0010791	-0.0003267	-0.0004946	-0.0034216	0.0002142	0.0014987
	(t-stat)	(0.09)	(-0.47)	(0.70)	(0.65)	(-0.20)	(-0.26)	(-1.37)	(0.09)	(0.60)
	R^2	0.9766	0.9740	0.9681	0.9670	0.9716	0.9549	0.9418	0.9407	0.9446
200301:200512	α_{APB}	0.000146	-0.0003178	0.0011283	0.0012328	-0.0006225	0.0006721	-0.0001819	0.0003338	0.0026247
	(t-stat)	(0.12)	(-0.26)	(0.81)	(0.85)	(-0.52)	(0.46)	(-0.09)	(0.13)	(1.23)
	R^2	0.9601	0.9583	0.9459	0.9492	0.9666	0.9485	0.9173	0.8891	0.9195
200401:200612	α_{APB}	-0.0013962*	-0.0020628***	-0.0008486	-0.0000574	-0.0019761*	-0.0022281*	-0.0024279	-0.0026567	0.0003739
	(t-stat)	(-2.02)	(-3.02)	(-0.92)	(-0.05)	(-1.85)	(-1.94)	(-1.24)	(-1.02)	(0.16)
	R^2	0.9800	0.9806	0.9665	0.9589	0.9606	0.9641	0.8895	0.8548	0.8768
200501:200712	α_{APB}	-0.0011827	-0.0011908	-0.0007198	0.0008691	-0.0004449	-0.0005439	-0.00171	-0.000278	0.0015947
	(t-stat)	(-1.47)	(-1.63)	(-0.69)	(0.85)	(-0.37)	(-0.43)	(-0.94)	(-0.11)	(0.61)
	R^2	0.9770	0.9804	0.9611	0.9694	0.9563	0.9585	0.9134	0.8817	0.8730
200601:200812	α_{APB}	-0.0007082	-0.0009602	0.0004213	0.0012815	0.0005674	0.0005834	-0.0049625**	-0.0015708	-0.0000144
	(t-stat)	(-0.58)	(-0.78)	(0.25)	(0.82)	(0.33)	(0.31)	(-2.49)	(-0.59)	(-0.00)
	R^2	0.9811	0.9823	0.9684	0.9741	0.9701	0.9650	0.9693	0.9461	0.9341
200701:200912	α_{APB}	-0.0011725	-0.0005138	0.0005225	0.0007315	0.0002548	0.0004897	0.0051846	0.0005323	0.0021775
	(t-stat)	(-0.79)	(-0.39)	(0.29)	(0.43)	(0.14)	(0.22)	(1.36)	(0.17)	(0.63)
	R^2	0.9787	0.9835	0.9691	0.9761	0.9729	0.9556	0.9240	0.9355	0.9170
200801:201012	α_{APB}	-0.0013598	-0.0011557	-0.0009771	0.0004643	0.0003774	0.0004662	0.0044131	0.0021241	0.0028341
	(t-stat)	(-0.88)	(-0.79)	(-0.54)	(0.25)	(0.20)	(0.19)	(1.22)	(0.71)	(0.85)
	R^2	0.9801	0.9824	0.9714	0.9742	0.9734	0.9527	0.9313	0.9445	0.9244
200901:201112	α_{APB}	-0.0008389	-0.00041	-0.0006945	-0.0001419	0.0010155	0.0017603	0.0067627**	0.0064439***	0.0061148**
	(t-stat)	(-0.45)	(-0.23)	(-0.35)	(-0.07)	(0.57)	(0.73)	(2.23)	(2.73)	(2.08)
	R^2	0.9582	0.9601	0.9456	0.9548	0.9610	0.9253	0.9029	0.9343	0.8955
201001:201212	α_{APB}	0.0005657	0.0001674	-0.000595	-0.000088	0.0003978	0.0014628	-0.0001397	0.0040205*	0.0023028
	(t-stat)	(0.31)	(0.10)	(-0.28)	(-0.04)	(0.21)	(0.59)	(-0.07)	(1.73)	(0.86)
	R^2	0.9469	0.9509	0.9258	0.9304	0.9422	0.9090	0.9258	0.9162	0.8911
201101:201312	α_{APB}	0.0021119	0.0010403	0.0000467	0.0007306	0.0011174	0.0004506	0.0034577*	0.0042219*	0.0023897
	(t-stat)	(1.10)	(0.56)	(0.02)	(0.32)	(0.56)	(0.19)	(1.78)	(1.69)	(0.86)
	R^2	0.9273	0.9306	0.9079	0.9120	0.9261	0.8994	0.9096	0.8761	0.8508
201201:201412	α_{APB}	0.0004278	-0.0004545	-0.0012847	-0.000187	-0.0010089	-0.0028762	0.0050169*	0.0003987	-0.0000386
	(t-stat)	(0.19)	(-0.20)	(-0.55)	(-0.08)	(-0.41)	(-1.01)	(1.71)	(0.13)	(-0.01)
	R^2	0.8977	0.8880	0.8823	0.8854	0.8796	0.8544	0.7684	0.7841	0.7676
201301:201512	α_{APB}	0.0004401	0.0006989	0.0010435	0.0006611	0.0011828	-0.000168	0.0053973	0.0015759	0.0039938
	(t-stat)	(0.16)	(0.25)	(0.38)	(0.24)	(0.41)	(-0.05)	(1.64)	(0.45)	(1.06)
	R^2	0.8780	0.8644	0.8590	0.8649	0.8448	0.8252	0.7670	0.7577	0.7000
Total (no	$\boldsymbol{\alpha_{APB}}$	**-0.0003346**	**-0.0004435**	**0.0003429**	**0.0004415**	**-0.0004291**	**-0.0001765**	**0.0002056**	**0.0011993**	**0.00107**
overlapping)	**(t-stat)**	**(-0.60)**	**(-0.86)**	**(0.58)**	**(0.64)**	**(-0.70)**	**(-0.21)**	**(0.20)**	**(1.14)**	**(1.05)**
	$\mathbf{R^2}$	**0.9526**	**0.9583**	**0.9447**	**0.9335**	**0.9397**	**0.9104**	**0.8837**	**0.8768**	**0.8870**

5.3.2. Fund Performance: Carhart vs. APB alphas

In this section, we focus on the comparison between UK fund performance obtained deploying standard Carhart model (equation 1) and the Hunter et al. (2014) APB model (equation 4). Figure 1 presents the difference in APB alpha and Carhart alpha, averaged across the funds (and peer-groups) in each of the 21 rolling periods. The difference in alphas is given in annualised basis points. The impact of non-zero APB alpha is as follows: if APB alpha is positive, in the periods when standard Carhart alpha for a fund is positive (negative), introducing the APB adjusted model will push the fund alpha downwards (upwards). In Figure 1 we document about 55% of time periods in which on average the APB alpha is higher than the standard Carhart one.

Figure 1: The average annualized difference (in bps) in monthly APB adjusted and Carhart alphas

Results are presented for each rolling period across all funds.

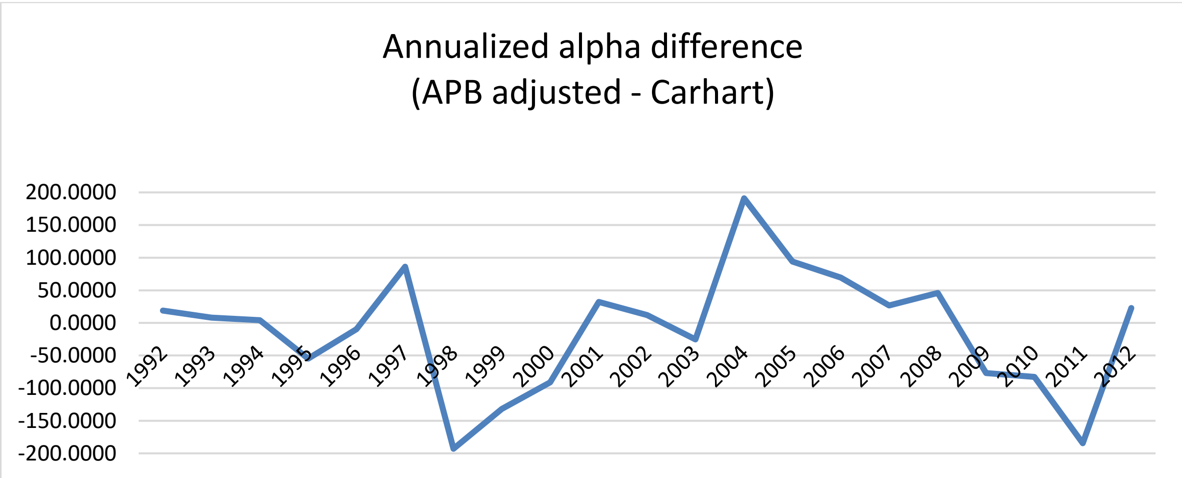
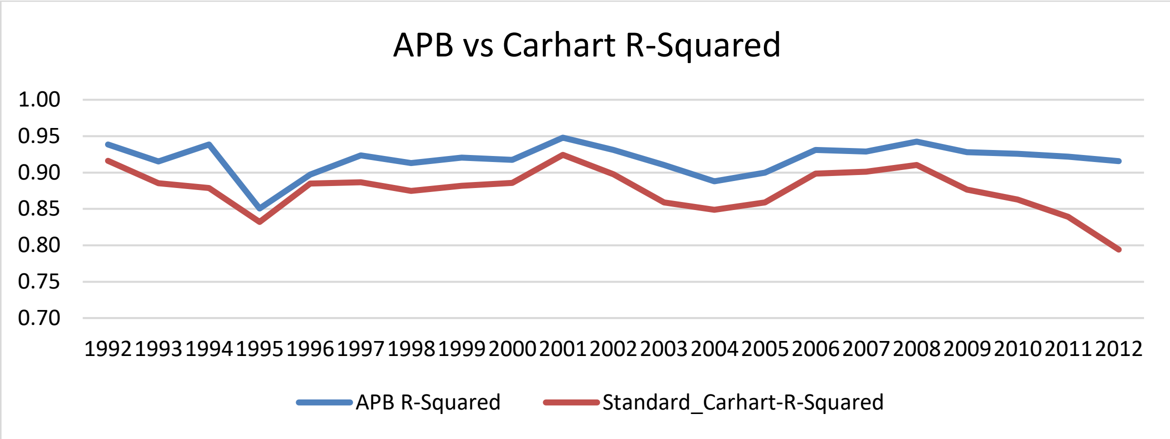


Figure 2 shows the average R-squared from the APB model and standard Carhart model. The average is given for all funds for each of the 21 rolling periods. It is evident that adding commonalities in fund strategies to the standard benchmark model factors results in greater explanatory power of the APB adjusted model, which is in line with Hunter et al. (2014).

Figure 2: The average R-squared (APB adjusted model (equation 4) vs Carhart model (equation 1))

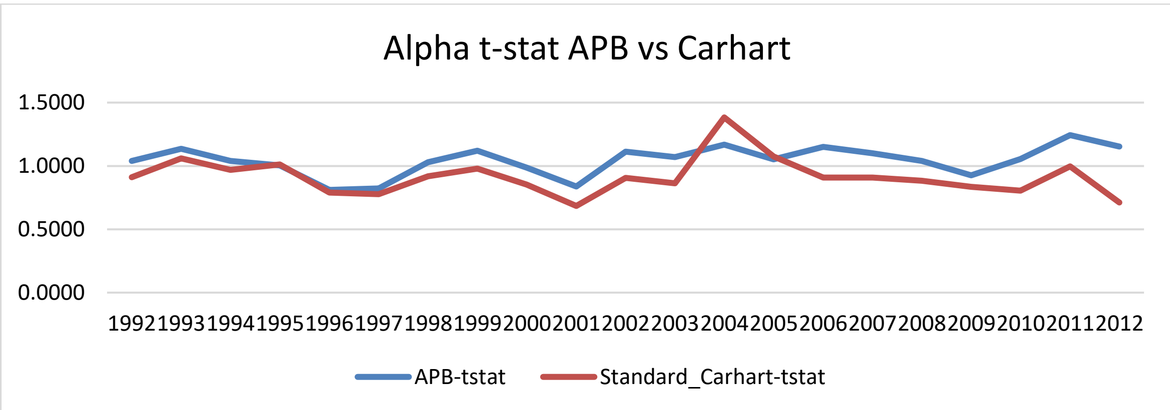
Results are presented for each rolling period across funds.



The information we obtain from Figure 1 and Figure 2 tells us that the APB adjusted model is more accurate when it comes to capturing funds' idiosyncratic risks. In addition, the APB adjusted model provides alphas that are of higher statistical significance than the standard Carhart alphas. This is true for all periods, except 2002-2004, as shown in Figure 3. The figure displays average absolute value of t-statistics of standard Carhart alpha and APB adjusted alpha across all the funds.

Figure 3: The average absolute value of t-statistics (standard Carhart alpha vs APB adjusted alpha)

Results are presented across all the funds in each rolling period.



Let us now look at the number funds that generate positive or negative APB adjusted alphas in each of the rolling periods. For each peer group of funds and each rolling period Table 3 presents the number of funds with positive/negative APB adjusted alphas in the first row and the number of funds with significantly positive and negative APB adjusted alphas in the second row. For the latter, we consider all the funds with significant alphas at 10% level or better. The third row shows the R-squared of the model in each instance. All regressions are performed per fund and per time period.

Similar to Hunter et al. (2014), when we control for the alphas earned by taking common risks in the peer group, there is approximately even split between positive and negative alphas. The percentage of significantly positive and negative alphas is approximately equal across peer-group categories and does not exceed 12.5% of the total number of funds in category/period in total.

In this section, we have established that the APB model provides investors with alphas showing unique fund managers skill, stripped out of the any common skill shared with other managers in the same peer group. As the number of funds with significantly positive alphas is proportionally small, the main question of this paper is to determine whether APB model is successful in picking funds with persistently positive and significant alphas, i.e. the funds that are persistently highly ranked.

Table 3: The number of mutual funds with positive/negative APB adjusted alpha

The table reports the number of funds with positive/negative APB adjusted alphas (the first set of numbers in each period) and number of significant positive/negative APB adjusted alphas (the second set of numbers in each period) from the following model estimated for each fund and each period from the table: $R_{i,t,u-R_{F,t,u}} = \alpha_{i,ADJ} + \beta_{i,M,t,u}\left(R_{M,t,u} - R_{F,t,u}\right) + \beta_{i,SMB}SMB_{t,u} + \beta_{i,HML}HML_{t,u} + \beta_{i,WML}WML_{t,u} + \beta_{i,ADJ}\left(\widehat{a_{APB,t,u}} + \widehat{e_{APB,t,u}}\right) + \omega_{i,ADJ,t,u}$. In the model $R_{i,t,u-R_{F,t,u}}$ is the excess return on fund I for period u (t is the frequency of the data, months and u represents the estimated subperiods in months); $R_{F,t,u}$ is the UK risk free rate; $R_{M,t,u} - R_{F,t,u}$ is the UK market risk premium; SMB and HML are Fama-French size and value factors and WML is the Carhart (1997) momentum factor. $\widehat{a_{APB,t,u}}$ is the alpha of the APB and $\widehat{e_{APB,t,u}}$ is the APB residual. $\left(\widehat{a_{APB,t,u}} + \widehat{e_{APB,t,u}}\right)$ is the adjustment factor in the APB-adjusted model and the $\alpha_{i,ADJ}$ is the APB-adjusted alpha. Numbers in **bold** mark the total across style categories (last column) and total sample period (last row).

Period	**Number of Mutual Funds Positive/Negative and Significant Positive /Significant Negative APB adjusted alphas**									
	Large Value	**Large Blend**	**Large Growth**	**Mid Value**	**Mid Blend**	**Mid Growth**	**Small Value**	**Small Blend**	**Small Growth**	**TOTAL**
199201:199412	31/29	30/39	8/6	6/6	5/6	2/1	2/1	4/3	9/9	**97/100**
	5/8	5/6	2/2	0/1	0/1	0/1	0/0	1/2	2/2	**15/23**
199301:199512	31/33	33/42	7/8	5/7	5/6	2/1	1/2	4/5	8/10	**96/114**
	5/8	8/11	3/3	1/3	1/0	0/0	1/1	1/2	1/2	**21/30**
199401:199612	32/32	30/49	8/8	6/6	3/8	1/3	1/2	5/4	10/9	**96/121**
	5/7	9/6	1/2	1/2	2/1	1/1	1/0	1/2	2/3	**23/24**
199501:199712	47/19	49/44	10/8	8/4	5/7	4/2	1/3	3/7	11/9	**138/103**
	13/1	7/6	3/1	1/0	2/1	1/1	1/0	1/1	1/1	**30/12**
199601:199812	38/33	52/46	12/6	7/5	6/6	4/2	3/1	5/5	12/10	**139/114**
	3/1	6/4	2/0	1/0	0/0	1/0	1/1	1/1	0/2	**15/9**
199701:199912	35/41	47/59	6/12	5/7	7/5	3/3	3/3	5/5	12/12	**123/147**
	2/2	8/10	1/2	0/0	1/0	0/0	0/1	0/2	0/3	**12/20**
199801:200012	33/51	50/63	8/10	4/8	8/7	5/3	4/2	4/8	13/12	**129/164**
	6/4	6/10	0/2	1/6	0/1	1/0	0/2	1/2	1/3	**16/30**
199901:200112	40/51	53/72	8/10	4/8	9/11	3/5	2/4	6/7	14/15	**139/183**
	4/6	13/15	2/5	2/6	1/4	2/2	1/1	1/2	3/3	**29/44**
200001:200212	44/54	67/76	11/8	5/8	11/12	3/6	3/3	6/9	13/17	**163/193**
	9/6	9/14	0/1	0/1	1/2	0/0	0/1	1/0	6/4	**26/29**
200101:200312	59/43	77/73	12/11	8/5	10/14	4/6	3/3	7/8	14/18	**194/181**
	6/4	9/4	1/0	0/1	1/1	1/0	1/0	2/2	3/4	**24/16**
200201:200412	56/56	86/92	14/13	8/7	16/15	8/3	2/4	8/9	17/17	**215/216**
	13/7	21/22	4/4	1/1	3/1	2/2	1/1	3/3	3/6	**51/47**
200301:200512	68/56	100/95	12/17	7/8	16/16	7/4	2/5	12/10	15/22	**239/233**
	4/12	14/23	2/5	1/1	3/0	1/1	1/0	2/3	7/6	**35/51**
200401:200612	70/67	107/105	16/14	7/11	14/22	6/7	4/4	13/12	19/19	**256/261**
	17/20	19/32	4/7	2/2	3/2	2/1	1/1	2/2	10/5	**60/72**
200501:200712	80/65	107/104	16/15	10/8	15/21	6/8	3/4	11/13	23/19	**271/257**

	11/16	13/25	4/3	1/1	3/1	1/2	1/1	3/1	6/4	**43/54**
200601:200812	89/62	124/101	16/19	13/8	21/19	7/10	5/3	12/13	24/22	**311/257**
	29/15	25/24	5/4	0/3	5/4	4/2	0/0	5/2	7/4	**80/58**
200701:200912	87/69	111/102	17/21	11/12	21/24	12/10	2/5	15/10	24/25	**300/278**
	23/12	29/22	3/0	5/5	7/5	3/2	1/0	0/1	4/2	**75/49**
200801:201012	83/69	98/93	18/19	9/12	19/25	14/7	3/4	10/12	29/20	**283/261**
	18/13	18/12	3/0	2/4	6/8	4/4	0/2	3/1	2/7	**56/51**
200901:201112	81/76	82/103	13/20	11/10	21/23	10/11	3/5	11/15	24/24	**256/287**
	12/10	17/15	2/1	0/1	4/7	1/0	1/0	3/2	5/4	**45/40**
201001:201212	77/69	75/98	12/14	9/11	24/16	10/10	4/4	11/14	25/18	**247/254**
	15/13	20/17	1/5	1/0	7/8	1/0	1/0	3/2	5/5	**54/50**
201101:201312	72/65	83/83	12/14	11/10	22/13	13/8	2/3	12/13	24/20	**251/229**
	20/20	23/22	5/3	2/2	5/7	3/1	0/0	4/2	10/8	**72/65**
201201:201412	70/66	79/66	13/12	12/8	18/13	11/9	3/3	14/13	20/25	**240/215**
	12/16	15/13	2/3	3/2	2/3	1/0	1/3	5/7	7/4	**48/51**
Total	**1,223/1,106**	**1,540/1,605**	**249/265**	**166/169**	**276/289**	**135/119**	**56/68**	**178/195**	**360/352**	**4,183/4,168**
	232/201	**294/313**	**50/53**	**25/42**	**57/57**	**30/20**	**14/15**	**43/42**	**85/82**	**830/825**

5.4. Predictive Ability of APB Adjusted Alphas

5.4.1. Ability of APB adjusted alphas to predict performance one year ahead

To test the predictive ability of APB adjusted alphas, we split the funds into quartiles according to their historical performance and assess their performance in the subsequent year, where Quartile 1 represents the best performing funds and so on. More precisely, the performance quartiles in year t are formed using t-test values of mutual fund APB adjusted alphas estimated using the information for the period t-36 months (equivalent to one rolling window in our study). If APB adjusted alphas themselves are used, it would be possible to have funds with positive yet insignificant alphas in Quartile 1 that would rank higher than some funds with significantly positive APB adjusted alphas. Ranking the funds by the values of their APB adjusted alpha t-test ensures that all the funds with significant positive (negative) APB alphas are placed in the top (bottom) quartile as the best (worst) performing funds. We require a minimum of 8 funds per category in each rolling period of t-36 months to form the quartiles[33].

To assess the predictive power of APB alphas to indicate the best (worst) performing funds we examine fund performance 12 months post period t (t+12m). For a fund to be included in the quartile allocation, we require that it has minimum of 30 months of returns data in each t-36 months rolling window over which the APB alpha is estimated. We also require that same fund to have minimum of 6 months of returns post quartile allocation, i.e. post period t (t+12m). This reduces the number of funds in the sample in this section of the paper to 748 (from 780 reported in Section 2).

The fund performance one-year-ahead, over period t+12m is gauged with fund's excess returns and Carhart alphas. Table 4 reports the difference in the excess return (first reported number) one-year-ahead and the four-factor alpha (second number) one year ahead between the top and the bottom quartile. Both excess returns and alphas are annualised values expressed in percentages. These differences are reported for each Morningstar style category (peer group), each of the 21 rolling periods[34] and in total: across the categories (last column) and the periods (last row). We use the z-test to determine significance of differences in performance between the top and the bottom quartile.

[33] The Small-Cap Value category fails this requirement in all the rolling periods, except 2009-2011, when only 8 funds are present. For this reason we chose not to show the results for that category. They are available on request.
[34] Note that in the Mid-Growth category due to insufficient number of funds it was not possible to create quartiles in the first 6 rolling periods.

Table 4: Difference in top and bottom quartile: one-year ahead Excess Returns/Carhart alphas (annualized, in %)

The funds are grouped into performance quartiles using t-statistics of APB adjusted alphas estimated using equation (4) and 36 months of historical data. One-year ahead performance is gauged through excess returns and Carhart alphas. The table shows the difference in one-year-ahead excess returns/Carhart alphas between two extreme performance quartiles: quartile 1(top) and quartile 4 (bottom). The difference is annualizes, in percent (%). *, **, and *** denote that Z-test for significance in the difference is significant at 10%, 5% and 1% levels respectively. Numbers in **bold** mark the total across style categories (last column) and total sample period (last row).

	Difference between Quartiles 1 and 4; Excess Returns/Carhart Alpha (annualized, in %)								
	Peer-Group Category								
Years	**Large Value**	**Large Blend**	**Large Growth**	**Mid Value**	**Mid Blend**	**Mid Growth**	**Small Blend**	**Small Growth**	**Total**
199201:199412	2.87/5.83	1.58***/3.89**	2.97***/3.03	0.11**/5.30	-2.53*/3.17	---/---	---/---	1.61**/-0.27	**1.70/4.04** **
199301:199512	1.80/1.48	0.84***/0.40	-6.40***/2.68	3.80/-2.85**	8.81**/7.81**	---/---	0.84**/1.27	4.73*/4.22	**1.84/2.57** **
199401:199612	-2.13*/2.84***	-0.36***/3.01	-0.76***/1.06	-7.29***/7.25***	8.88/6.43	---/---	0.20/4.00	-3.03/-2.57	**-1.25/3.21*** **
199501:199712	-4.20***/0.15	-3.95***/2.94	-7.44/-3.89	1.82/7.94	-8.57/-7.56***	---/---	-8.29/-0.36	-7.02/-1.71	**-2.09/0.42***
199601:199812	-1.56/3.48	5.77***/2.10**	1.42**/0.24	41.43/8.88	21.90/9.50	---/---	2.53***/27.55	9.00***/3.62	**4.55/3.27**
199701:199912	-1.18/4.79**	2.54***/4.68	9.44***/4.79	19.73/14.97	-8.38/-2.51	---/---	6.76/-5.64***	3.49/6.05	**2.95*/5.31*** **
199801:200012	3.03/-0.29	4.32***/1.93	10.36***/6.31	11.53/-1.10	7.56/9.43	0.53/5.07	12.48***/6.79	-0.34***/2.82	**4.69***/2.04**
199901:200112	-1.41/0.30	-1.03***/-0.83***	-3.34***/-13.22	-3.04**/-7.02	-3.46***/-3.80	0.29***/6.46	5.73***/9.71	-4.72***/-10.94	**-1.62**/-2.02*** **
200001:200212	4.35***/-0.10**	3.09***/-1.44***	-0.97/1.55	9.57/-8.16	5.52/-1.13	-0.78/0.23	6.64*/17.41*	1.47/2.32	**4.06***/-0.34***
200101:200312	2.46**/3.47	2.22*/0.49	-3.98/-1.39	4.36*/2.33	-2.09/-3.70*	10.21/9.59	2.55/0.13	4.83*/2.41	**1.84*/1.50**
200201:200412	0.78/0.18	0.29/1.20	-0.61/4.30	1.53/-0.51	8.26/3.67	-2.46/-8.53	5.49*/-0.25	1.70/0.97***	**1.28**/0.67**
200301:200512	3.06***/2.87	1.56/2.65*	2.86/7.11	1.86/-1.50	8.05**/-1.43	19.93***/12.57	6.60/4.78	10.76***/9.66	**3.75***/3.32***
200401:200612	-1.91/-0.73	-3.21***/-2.68**	0.61/-0.93	-11.29***/-4.71	-3.08/2.73	-6.02/3.44	9.40***/9.80	5.20**/3.23***	**-2.71***/-0.49**
200501:200712	4.86***/3.45	-0.37/0.84	7.21***/4.85	-5.97/-0.70	6.35/3.62	11.21/16.53	4.02/18.29***	5.00**/12.63***	**2.51**/3.74*** **
200601:200812	-3.90/-3.02**	-2.26/-2.35	-15.74***/1.74	3.16/4.87	20.93**/10.45	16.53**/8.18	-18.43**/-0.22	5.62/-4.99***	**-2.74/-1.47**
200701:200912	1.74*/3.28***	4.74***/5.39	1.32/2.63	4.84/-0.48	8.83**/7.60*	14.14***/8.52	-3.21/1.31	6.87/-1.48	**3.99***/3.60** **
200801:201012	-1.64*/0.09***	0.20/0.71***	-4.41/-3.78	-1.43/-2.19	-1.56/0.96	10.54/10.38	-1.86/-2.40	5.14*/8.15**	**-0.39/0.74*** **
200901:201112	1.29/4.79***	2.57**/3.78**	0.74/6.54	-9.13**/-2.22	7.76**/7.31	5.45/1.20	8.39/17.53**	2.05/3.71	**3.13***/5.18*** **
201001:201212	6.28***/-0.13***	4.30***/6.83***	4.81/9.85**	12.34**/6.58	8.34/4.57	4.56/-2.98	4.92/9.31***	6.79/-3.68**	**5.61***/3.18*** **
201101:201312	3.30***/3.30***	3.63***/3.62**	2.92/3.07	7.19***/7.55	2.87/2.37***	0.91/0.86*	4.64*/5.50	-0.83/-0.36*	**3.05***/3.07*** **
201201:201412	-0.14/0.02	2.42/2.24	0.89/2.86***	0.78/0.50	3.07/2.09***	3.92/2.31	-0.37/9.76***	-0.68/0.77***	**1.37/2.38*** **
Total	**3.25***/1.28** **	**3.52***/1.47** **	**2.03/2.29**	**6.69**/1.48**	**7.44***/3.48** **	**4.09/3.32**	**0.80/6.88*** **	**3.87/1.43*** **	**3.55***/1.86*** **

For all peer groups and years, Table 4 shows that the difference in excess return between the top and bottom quartile is 3.55% p.a., while the difference in four factor alphas is 1.86% p.a. (both values significant at 1% level). In 15 (13) of the 21 rolling periods the four-factor alpha (excess return) differential is statistically significant. Reverting the attention to each peer-group separately, APB adjusted alpha has weaker predictive power of performance in Large-Growth and Mid-Cap-Growth categories, where the differences in performance one-year-ahead are still in favour of the top quartile, albeit not being statistically significant. This shows that, overall, by picking the funds with most significantly positive historical APB adjusted alphas will enable investor to generate higher excess returns and higher four-factor alphas in the subsequent one year period.

As a robustness test, we split the funds into deciles in period t, according to their APB alpha t-values, obtained for period t-36m. We follow the same steps as previously with quartiles and assess the ability of APB adjusted alpha to predict one-year-ahead performance by measuring the difference in excess return and four-factor alphas for two extreme deciles. In this case, due to the small number of funds in all except Large-Value and Large-Blend category, we perform this analysis for all peer-groups together per each period (equivalent to the last column of Table 4) and in total across the 21 rolling sub-periods. Table 5 reports the excess returns and four-factor alphas 12 months post decile formation for the top and bottom decile; their difference and the z-test for the significance of that difference. Across all the periods, top decile has higher excess return by 4.93% per year and higher alpha by 4.4% p.a. than the bottom decile (both significant at 1% level). Given that we do not separate performance by peer-groups, one may argue that the performance of the top (or bottom) decile can be driven by a particular Morningstar category that dominates the top or the bottom decile. Looking at percentage of funds from each peer-group in the top and bottom decile, we find that all peer groups are represented (approximately) equally: their weights ranging from 9.7% to 12.7% (8.8% to 13.7%) in the top (bottom) decile. Hence, regardless of the number of performance sets used in fund ranking, the investors selecting the funds from the top set will generate statistically and economically significantly higher excess returns/alphas in the following 12-monthly period.

Table 5: Difference in top and bottom decile: one-year ahead Excess Returns/Carhart alphas (annualized, in %)

The table shows results for funds across all styles grouped into deciles based on APB-adjusted alphas t-stats. The table reports one year-ahead excess returns/Carhart aphas for top and bottom decile, their difference, the z-test for the difference and the number of funds in top/bottom decile. The difference is annualises, in percent (%). *, **, and *** denote that z-test for significance in the difference is significant at 10%, 5% and 1% levels respectively. The numbers in **bold** correspond to the total sample period.

-

ALL STYLES	**Deciles Returns/alpha (annual)%**				
Period	**Top Decile**	**Bottom Decile**	**Difference (top-bottom)**	**Z-stat**	**# Funds**
199201:199412	16.90***/2.60*	12.52***/-1.11	4.39/3.72	2.21**/1.29	39
199301:199512	14.33***/3.66***	12.19***/-0.54	2.14/4.20	1.13/1.92*	42
199401:199612	11.76/8.06	15.02/3.68	-3.26/4.37	-1.19/4.29***	43
199501:199712	1.24/2.02*	6.94***/2.17	-5.70/-0.15	-2.22**/1.09	49
199601:199812	35.38***/-2.03	32.12***/-6.44**	3.26/4.41	0.41/-0.67	51
199701:199912	-1.03/5.63***	-6.35***/2.20	5.32/3.43	2.30**/2.19**	54
199801:200012	-13.60***/0.63	-16.18***/-0.35	2.58/0.97	1.59/0.36	59
199901:200112	-26.73***/-7.85***	-25.21***/-4.49***	-1.52/-3.36	-1.10/-3.08***	66
200001:200212	23.78***/1.90**	18.61***/2.31**	5.18/-0.42	2.84***/1.57	71
200101:200312	12.87***/-0.66	9.50***/-3.19***	3.37/2.53	2.14**/-0.52	75
200201:200412	17.91***/1.49*	14.84***/-0.58	3.07/2.07	2.73***/1.14	87
200301:200512	16.90***/-0.58	11.19***/-5.03***	5.71/4.45	3.91***/-0.58	96
200401:200612	-5.55***/0.37	-3.01***/-0.19	-2.54/0.57	-2.30**/0.38	98
200501:200712	-32.28***/4.25***	-34.39***/-0.98	2.11/5.24	1.37/2.62**	101
200601:200812	32.84**8/0.87	29.55***/-0.49	3.30/1.36	1.22/0.54	106
200701:200912	20.27***/-2.98**	16.55***/-5.32***	3.72/2.33	2.62**/-1.53	105
200801:201012	-4.70***/4.77***	-5.72***/2.22***	1.03/2.55	0.92/4.42***	101
200901:201112	18.70***/3.31***	15.92***/0.05***	2.77/3.26	1.69*/2.35**	99
201001:201212	33.49***/13.18***	28.09***/7.48***	5.41/5.70	2.93***/4.63***	91
201101:201312	-2.63***/-2.66***	-7.26***/-7.28***	4.64/4.62	4.58***/-2.69***	87
201201:201412	3.77***/7.99***	1.50/2.68***	2.27/5.32	1.27/6.01***	87
TOTAL	**7.46***/1.81*** **	**2.40***/-0.61** **	**5.06/2.42**	**4.93***/4.40*** **	**1,603**

To corroborate the findings from Table 4, that positive most significant historical APB adjusted alphas indicate a better performance one year ahead, we run the following cross-sectional regression model:

$$\alpha_{i,t,u+12} = a_i + b_i \alpha_{iADJ,t,u} + u_{i,t,u} \qquad (5)$$

Where $\alpha_{i,t,u+12}$ is the Carhart alpha of fund i one year ahead, i.e. 12 months following the estimation of APB adjusted alpha, $\alpha_{iADJ,t,u}$, in period *u*, using *u*-36 months of data. The model tests for persistence in performance in the cross section and it is run for each of the 21 rolling periods and a full sample period.

Table 6 lays out the results and reports the slope coefficient from equation (5), its t-statistics (in parentheses) and the R-squared of the model. The final column illustrates the impact that 100bp increase in APB adjusted alphas has on subsequent performance.

Table 6: Predictive ability of ABP adjusted alpha

The table reports the results from the equation (5): $\alpha_{i,t,u+12} = a_i + b_i\alpha_{iADJ,t,u} + u_{i,t,u}$; Where $\alpha_{i,t,u+12}$ is the Carhart alpha of fund i one year ahead, i.e. 12 months following the estimation of APB adjusted alpha (*t* is the frequency of the data, months and *u* represents the estimated subperiods in months), $\alpha_{iADJ,t,u}$, in period *u*, using *u*-36 months of data. The model tests for persistence in performance in the cross section and it is run for each of the 21 rolling periods and a full sample period. The numbers in **bold** correspond to the total sample period. Model parameters are in decimal points.

Period	Constant (a_i)	Beta (b_i)	R-Squared	Impact of 100 bp increase in $\alpha_{iADJ,t}$ on $\alpha_{i,t+12}$ (in bps)
199201:199412	0.0001705 (0.45)	0.3847532*** (3.56)	0.0611	39
199301:199512	-0.0003321 (-0.95)	0.2058412** (2.11)	0.0209	21
199401:199612	0.0039921*** (13.37)	0.3604046*** (3.67)	0.0588	36
199501:199712	0.0005548 (1.52)	0.0995939 (0.93)	0.0036	10
199601:199812	-0.0022935 (-3.73)	0.449866** (2.12)	0.0176	45
199701:199912	0.0024797*** (5.66)	0.5945969*** (3.79)	0.0508	60
199801:200012	0.0005741*** (1.69)	0.1312792 (1.36)	0.0063	13
199901:200112	-0.0032328*** (-7.49)	-0.3259747*** (-3.46)	0.0360	-33
200001:200212	0.0014164 (6.14)	-0.0878339 (-1.54)	0.0067	-9
200101:200312	-0.0010094*** (-3.73)	0.1087426 (1.33)	0.0047	11
200201:200412	-0.0002142 (-1.08)	0.0478412 (0.80)	0.0015	5
200301:200512	-0.0025381*** (-12.78)	0.3933201*** (5.30)	0.0566	39
200401:200612	-0.0000936 (-0.41)	-0.0483668 (-0.58)	0.0007	-5
200501:200712	0.0007838* (1.88)	0.4676845*** (2.72)	0.0146	47
200601:200812	0.0011985 (2.85)	-0.0271228 (-0.24)	0.0001	-3
200701:200912	-0.0028372*** (-7.85)	0.3106142*** (3.34)	0.0210	31
200801:201012	0.0023419*** (11.14)	0.0640666 (1.18)	0.0027	6
200901:201112	0.0011069*** (3.48)	0.4719138*** (4.24)	0.0352	47
201001:201212	0.0096663*** (19.15)	0.3540427** (1.97)	0.0084	35
201101:201312	-0.0040424*** (-23.27)	0.3323759*** (6.22)	0.0825	33
201201:201412	0.0043385*** (17.65)	0.344805*** (4.30)	0.0408	34
TOTAL	**0.0006119*** **	**0.1552812*** **	**0.0047**	**15**

	(7.29)	**(6.14)**		

The results overall illustrate a statistically strong positive relation between APB adjusted alphas based on historical returns and future four-factor alphas. In the total sample, a fund with a 1% increase in APB adjusted alpha will have 15.53bps higher four-factor alpha one year ahead. One might argue that the magnitude of increase in alphas is not large, but it should be taken into consideration that the average monthly APB alpha in the total sample is 11bps. Hence an increase of 1% on 11bps alpha will raise Carhart alpha by 15.53 bps one-year-ahead.[35] What is certain here is that this positive relationship between historical APB adjusted and future four-factor alpha is significant at 1% level in the total sample period and is present in 17 out of 21 rolling sub-periods (significant at least at 5% level in 13 of those).

5.4.2. Ability of APB adjusted alphas to predict funds' future quartile rankings

We have concluded from Section 5.4.1. that higher APB adjusted alpha signifies better fund performance one-year-ahead. In this section, we utilize the ordered probit model to determine the probability that a fund will move up the performance quartile ranks one-year-ahead, if the significance of its APB adjusted alpha increases[36]. To this end, the funds are ordered in quartiles according to their Carhart alphas in year u+1. The ordered probit model is given as:

$$y_i^* = x_i \beta + e_i \qquad (6)$$

Where the dependent variable (y_i) is observed and represents ordered outcomes, in our case quartile ranks: rank 0 is assigned is a fund i is ranked in the bottom quartile, rank 1 if in quartile 2, rank 2 if in quartile 3 and rank 3 if the fund is in the top quartile based on its Carhart alpha in year u+1. The explanatory variable x_i is the value of the t-test corresponding to fund's ABP adjusted alpha estimated for the previous year, year u, estimated using the 36 months of

[35] In the similar manner we have tested the predictive ability of the standard Carhart alpha and obtained that the overall marginal effect (beta from equation 5) was reduced by 58% when historical Carhart alpha is used, from 15.53 to 9.8bps; the difference being significant at 5% level.

[36] The ordered probit model is used as a robustness check to access whether t stat in period u-36 to u is a good predictor of performance one year ahead (u+12). Mutual funds are ordered regarding their alpha quartile one year ahead among Quartile 1 (top quartile) and Quartile 4 (bottom quartile). These are observed categories. What is aimed to address is whether APB alpha t-stat in the previous 36 months is a good predictor for the mutual fund quartile location one year ahead. We do not aim to analyse whether funds will move in between quartiles. The aim is to access whether t-stat is a good predictor of the mutual fund "location" one year ahead.

historical data. e_i is the independent and identically distributed random variable. The observed dependent y_i is determined from y_i^* and follows the following conditions:

$$y_i = j \quad \text{if} \quad \gamma_{j-1} \leq y_i^* \leq \gamma_j \tag{7}$$

The threshold values gammas, γ, are estimated along with the β coefficient using the maximum likelihood estimation. The value of the observed variable y_i depends on whether or not the gamma thresholds have been crossed. Therefore, in order to evaluate the probabilities of observing each value of y_i, the following calculations are required:

$$P\,(y_i = j|\, x_i, \beta, \gamma) = P\,(\gamma_{j-1} \leq y_t^* \leq \gamma_j) = F(\gamma_j - x_i\beta) - F(\gamma_{j-1} - x_i\beta) \tag{8}$$

Table 7: Ordered Probit – Measuring ability of APB adjusted alphas to predict funds' future quartile rankings

The table reports model coefficient, the maximum alpha threshold per quartile, the number of observations, Wald Chi-squared statistics and Marginal Effects of the ordered probit model from equation (6) and described in section 4.2.: $y_i^* = x_i\ \beta + \varepsilon_i$ Where the dependent variable (y_i) is observed and represents ordered outcomes, in our case quartile ranks: rank 0 is assigned is a fund i is ranked in the bottom quartile, rank 1 if in quartile 2, rank 2 if in quartile 3 and rank 3 if the fund is in the top quartile based on its Carhart alpha in year *u*+1. The explanatory variable x_i is the value of the t-test corresponding to fund's ABP adjusted alpha estimated for the previous year, year *u*, estimated using the 36 months of historical data. The numbers in parentheses are z-statistics and *, ** and *** denotes significance at 10%, 5% and 1% level.

Max alpha threshold (annualized, in %)	**Top Quartile**	11.94%	**Marginal Effects**	**Top Quartile**	0.0237*** (8.37)
	Quartile 2	2.44%		**Quartile 2**	0.0061*** (7.88)
	Quartile 3	-1.69%		**Quartile 3**	-0.0061*** (-7.88)
	Bottom Quartile	-8.67%		**Bottom Quartile**	-0.0236*** (-8.37)
Model Coefficient (β) (t-test)			0.0746*** (8.38)		
#Obs.			8,007		
Wald Chi-sq.			70.51		

Table 7 lays out the results. We report the maximum annualized alpha threshold points among bottom and third quartile -8.67%), third and second quartile (-1.69%) and third and top quartile (2.44%). Therefore to be located in top quartile one year ahead a fund will have to have alpha of at least 2.44% per year.

What is of greatest interest to investor in this table are the marginal effects from the ordered probit model given in the last column. They represent the increase/decrease in probability that a fund will move to the top quartile due increase in t-stat alpha in the previous 36 months.

Hence, a one unit increase in t-statistics of APB adjusted alpha in period t (using t-36 months of data) will increase probability that that fund will appear in the top quartile by 2.37% and decrease probability that a fund is classified in the bottom quartile by 2.36%.

On the whole, the results in this section complement the general persistence results from section 5.4.1. We show not only that a fund with a greater APB adjusted alpha t-test will have better future performance, but that small increases in APB adjusted alphas t-test significantly improve investors' chances of holding a portfolio in the top performance quartile one-year-ahead.

5.4.3. Is the performance persistence driven by winner or loser funds?

We have established so far that persistence when performance is assessed with APB adjusted alphas in UK funds exists, but it is not clear whether it is more prominent among loser funds, as the previous literature suggests. To answer this question, we adapt Fletcher and Forbes (2002) approach of contingency tables and get insight into the persistence in performance by fund category.

To form contingency tables, we differentiate between four groups of funds according to their performance in two consecutive periods (years): winner/winner group (W/W) are funds whose APB adjusted alpha t-test was in the top quartile in period one[37] and their performance one year ahead (in period two) remains above the median (i.e. in quartile one or quartile two); winner/loser (WL) group are those that were winners in period one and losers in period two; loser/winner (L/W) are the opposite of W/L; and the loser/loser (L/L) funds are having the lowest 25% of APB adjusted alpha t-stats in period one and performance below the median one year ahead. For robustness, we are using three different measures of performance one-year-ahead: Carhart alphas, t-test of Carhart alphas and excess returns. The number of funds in each of WW, WL, LW and LL groups are then counted in each rolling sub-period and aggregated in total, over the whole sample period.

[37] Our results remain robust when the funds are split into winners and losers according to the values of APB adjusted alpha, not its t-statistics. Note that four quartiles are formed as before: according t-statistics associated with APB adjusted alphas estimated for a fund for period t, using t-36 months of historical data.

To test for significance in persistence and get insight into the drivers of persistence, we apply the Brown and Goetzmann (1995) log-odds ratio approach:

$$\text{Log} - \text{odds ratio} = \ln \frac{\text{WWxLL}}{\text{WLxL}W} \quad (9)$$

The standard error of the log-odds ratio is given as:

$$\text{SE}_{\text{log-odds}} = \sqrt{\left(\frac{1}{\text{WW}} + \frac{1}{\text{WL}} + \frac{1}{\text{LW}} + \frac{1}{\text{LL}}\right)} \quad (10)$$

Table 8 presents the log-odds ratios and their significance. Panel A is based on Carhart alpha t-stats, Panel B on Carhart alphas and Panel C on excess returns as measures of performance one-year-ahead. The results presented are aggregate results for the total sample period, 1994-2016.

Table 8: Contingency Table for Persistence in Performance, by fund peer-group Morningstar style category

The table reports the number of winner/winner, winner/loser, loser/winner and loser/loser funds for each Morningstar peer-group; the log-odds ratio, Chi-squared test and the number of funds per category. Significance of log-odds ratios is given by z-statistics and *, ** and *** denotes significance at 10%, 5% and 1% level. Panel A shows performance based on t-test of Carhart alphas one year ahead, Panel B on Carhart alpha and Panel C on excess returns.

Panel A	**Carhart t-test**				# Funds
Large Value	Winner	Loser	Odd-ratio	Chi-Squared	
Winner	327	241	1.66***	17.91***	1,120
Loser	248	304			
Large Blend					
Winner	436	320	1.64***	22.35***	1,487
Loser	332	339			
Large Growth					
Winner	76	54	1.81**	5.21**	242
Loser	49	63			
Mid Value					
Winner	50	37	1.22	0.38	163
Loser	40	36			
Mid Blend					
Winner	83	56	2.00***	7.60***	261
Loser	52	70			
Mid Growth					
Winner	32	28	2.03*	3.30*	110
Loser	18	32			

Small Blend					
Winner	66	30	4.75***	23.90***	175
Loser	25	54			
Small Growth					
Winner	106	77	1.77***	7.06***	350
Loser	73	94			
TOTAL					
Winner	1177	844	**1.75*** **	**75.82*** **	**3,912**
Loser	838	1053			

Panel B	**Carhart alpha**				# Funds
Large Value	Winner	Loser	Odd-ratio	Chi-_Square	
Winner	325	244	1.61***	15.69***	1,120
Loser	250	302			
Large Blend					
Winner	437	319	1.75***	28.70***	1,487
Loser	321	410			
Large Growth					
Winner	76	54	1.88***	5.86**	242
Loser	48	64			
Mid Value					
Winner	49	38	1.51	1.72	163
Loser	35	41			
Mid Blend					
Winner	80	59	1.77***	5.18**	261
Loser	53	69			
Mid Growth					
Winner	31	29	1.74	2.05	110
Loser	19	31			
Small Blend					
Winner	69	27	5.86***	29.97***	175
Loser	24	55			
Small Growth					
Winner	103	80	1.58**	4.52**	350
Loser	75	92			
TOTAL					
Winner	1171	851	**1.77*** **	**79.21*** **	**3,912**
Loser	826	1065			

Panel C	Excess Return				# Funds
Large Value	Winner	Loser	Odd-ratio	Chi-_Square	
Winner	343	226	1.93***	29.41***	1,120
Loser	243	309			
Large Blend					
Winner	424	332	1.60***	20.57***	1,487
Loser	324	407			
Large Growth					
Winner	72	58	1.72**	4.34**	242
Loser	47	65			
Mid Value					
Winner	47	40	1.80	3.45*	163
Loser	30	46			
Mid Blend					
Winner	82	57	1.81***	5.65**	261
Loser	54	68			
Mid Growth					
Winner	31	29	1.48	1.02	110
Loser	21	29			
Small Blend					
Winner	64	33	2.20***	6.98***	175
Loser	37	42			
Small Growth					
Winner	98	85	1.38	2.26	350
Loser	76	91			
TOTAL					
Winner	1162	861	**1.71*** **	**70.11*** **	**3,912**
Loser	833	1058			

The values of the odds ratio for the total period across all the fund categories are all positive and significant at 1% level, indicating strong persistence in performance. While all odds ratios in the table are above one, indicating persistence (as opposed to reversal) in performance, the Mid-Cap Value style category is the only one that does not exhibit significant persistence across the three performance measures of one-year-ahead performance. Also, Mid-Cap Growth category shows comparatively weak persistence, but all other Morningstar fund categories exhibit very strong persistence in performance, with Small-Blend category having the highest log-odds ratio of 4.75 when performance one-year-ahead is gauged with alpha t-stats, significant at 1%. The results per each of the 21 rolling periods and style category are available on request from authors. They are in line with the aggregate results, overwhelmingly showing (in around 75% of periods across different styles) the odds ratio above one, indicating persistence in performance within fund categories. We also standardize the results by adjusting the number of funds per style in each rolling window by their average value[38], the results do not change showing that they are not driven by rolling periods with more funds.

In all style categories there is more funds in winner/winner and loser/loser fund groups than those in the intermediate two groups. Hence, the performance persistence is stemming both from winners and losers. With the exception on Mid-Cap Growth style category, all styles exhibit marginally higher number of winner/winner funds than loser/loser funds implying that good performance is marginally more likely to repeat than the poor one. This is at odds with a number of studies on persistence in performance that are based on standard factor models, who find more persistence among loser funds.

5.5. Conclusions

We contribute to the literature of UK equity mutual fund performance and persistence in performance, from the perspective of a peer group. We use approach by Hunter, Kandel, Kandel and Wermers (2014) that enables us to identify the top performers within each peer group by accounting for idiosyncratic risks common to all funds within a peer group. This is the first study to apply the peer-group adjusted alpha method for performance evaluation in the UK. We also test for persistence in performance one-year-ahead by assessing whether mutual funds with the highest adjusted alphas within a peer group will continue to be the top performers one year later.

Our sample is comprised of 817 funds over the period January 1992 to February 2016. The funds are split into nine Morningstar categories (3x3 combination of three size and three style categories), that we regard as the peer-groups. 65% of our funds reports FTSE All Share Index as their benchmark even though less than half (310) of our funds are in the Large cap Blend Morningstar category, for which this index could be considered an appropriate benchmark, although FTSE 100 may be a better fit. Active Peer-group Benchmark (APB) approach proposed by Hunter et al. (2014) avoids this problem of inadequate benchmarks and calculates APB return as the equally weighted return of all the funds in the same Morningstar category. They modify the standard Carhart four factor model by adding APB's four factor alpha and the error term. The new model is then enabling us to identify funds that exhibit performance above that earned by the average skilled manager in the group. We find that APB adjusted model has

[38] Results available on request.

higher R-squared and that alphas from the model are more statistically significant compared to the standard Carhart model.

In assessing persistence, we form four performance quartiles based on historical APB adjusted alphas' t-test and evaluate performance of funds one-year-ahead using funds' excess returns and Carhart alphas. We test persistence overall and by fund peer-group using both parametric (regression) and non-parametric (contingency tables) method. The performance is found to persist regardless of the method employed and results remain robust when funds are split into deciles rather than quartiles. We conclude that APB adjusted alphas have strong predictive ability of future returns and that 1% increase in t-statistics of APB-adjusted alphas leads to 2.37% increase in probability that a fund will be placed in the top performing quartile. Our findings reveal that persistence is driven by both winner and loser funds, contrary to existing evidence from the UK attributing persistence mainly to poor performers. The result is consistent across Morningstar peer-groups.

This study of relevance to academics and both individual and institutional investors as it illustrates how APB adjusted alpha approach can be used to identify funds with superior relative performance within a peer-group. In the scope of revision of benchmark selection and reporting by funds, policy makers could request that the fund identifies its peer-group and that performance relative to that peer group is made available to investors. The study can be modified and extended to other types of funds (in the same and different asset classes) where benchmarking is ambiguous.

Chapter 6

Conclusions

Due to the mutual fund's important role in the investment activities of individual investors the performance of mutual fund/fund managers has been extensively examined in finance literature. The US evidence on fund performance is controversial. A vast share of academic studies claims that mutual funds generate excess performance that is just enough to cover their expenses. However, some studies such as Cremers and Petajisto (2009) and Hunter et al. (2014) which propose new methodologies with the aim to examine the actual type of active fund management and estimate fund relative performance versus peers, posit that funds are able to outperform their benchmarks and show some performance persistence. The evidence in the UK is scarce and mainly documents the underperformance of UK unit trusts.

Fund risk-adjusted performance has generally been estimated with the use of Fama and French three-factor and the Carhart four-factor-models. Both of the models have been widely accepted by academics and practitioners with the last model being considered more accurate. It is a common practice among investors to make judgements on fund performance by comparing the excess return of a fund to the return of a passive benchmark with the same risk characteristics. Several studies (for instance Angelidis et al.,2013) emphasized the importance of considering the fund self-reported benchmark for more precise inferences on fund manager performance. However, recent literature indicated potential problems in the previous evidence on mutual fund performance and performance persistence. Thus, Cremers, Petajisto and Zitzewitz (2012) provide evidence of non-zero alphas obtained for passive benchmark indices when regressed against risk factors of standard Fama and French and Carhart models. As a consequence, this evidence indicates that the inferences on mutual fund performance may be significantly biased if the alphas estimated with the standard benchmark models were not adjusted for the positive/negative benchmark alphas. To continue, Sensoy (2009) provide evidence that mutual fund self-declared benchmarks, which are commonly used as a reference benchmarks in fund performance evaluation, are constantly mismatched for 31.2% of funds. As a result, fund performance estimated versus inaccurate benchmark may lead to wrong conclusions and investment decisions.

Considering the above, the aim of this dissertation entitled *Benchmark Indices, Alpha creation and performance persistence* is threefold: 1) to revisit the performance of mutual funds in the UK (using the Angelidis et al. (2013) methodology hitherto tested in the US), 2) to measure the impact of inaccurate benchmark selection and the extend of possible biases in fund performance assessment (US market); 3) estimate relative fund performance and performance persistence in the UK accounted for non-zero benchmark alphas.

To conduct this analysis we split this dissertation into three main sections (empirical essays) which we call chapter 3, chapter 4 and chapter 5. The rest of the dissertation includes chapter 1 introduction, chapter 2 the existing literature review in the area of research and chapter 6 conclusions.

The outcomes for the chapter 3 are based on the sample of 887 UK active funds for the period from January 1992 to October 2013. Similar to Cremers, Petajisto and Zitzewitz (2012) we document that standard Fama and French and Carhart models provides non-zero alphas for passive benchmark index FTSE 100. However, in contrast to the US the UK benchmark alphas are negative (-1.12% and 1.13%, with the three and four-factor models respectively, statistically significant at 1% level.). In addition, the results show that the UK benchmark alphas vary depending on the market conditions (from -1.61 and -2.86% during the bear market and from -0.47 to -1.10% for the bull market). Therefore, our results show that wrong factor/portfolio construction of the Fama and French and the Carhart model leads to amplified underperformance of UK equity mutual funds. When the performance of funds is adjusted for the negative alphas with the Angelidis et al. (2013) approach we show evidence that UK focused equity funds are able to generate positive excess returns (in contrast to the previous literature for the UK). To illustrate, the adjusted Fama and French alpha for the total period and the whole fund sample has dramatically increased from 13.81bps with the standard approach to 143.64bps per year when modified for the negative alpha in FTSE 100. A greater increase in alphas after the adjustment was observed in bear rather than in bull market periods.

To extend our analysis further we control the impact of negative benchmark alphas on the performance of funds grouped by investment style/strategy. The analysis is conducted for nine style categories in accordance with Morningstar database. Similarly to the previous results we document improved after-adjustment performance across fund investment styles. In addition, we show that over 70% of mutual funds concentrate their portfolios in Small/Value, Small/Growth and Small/Blend stocks. These funds perform better in comparison to other styles and deliver AGT-adjusted excess return of 1.62%, 2.04% and 1.54%, respectively; statistically significant at 1% level). The performance persists even during market downturns. Overall, Small/Value style funds showed the most consistent outperformance and Large/Value funds provided the best performance versus others during the financial crisis 2008-2009. Robustness test for the choice of benchmark index revealed that when FTSE 100 benchmark is replaced with style-specific FTSE Small Cap Index for small cap funds the results supported our previous inferences and became even more evident.

In the chapter 4 we analysed the impact of benchmark choice on US equity funds performance and estimated to which extend inaccurate benchmark selection affects absolute and relative performance evaluation. For the analysis we utilised a sample of net monthly returns of 1281 actively managed US equity mutual funds from January 1992 to February 2016. In accordance with the Morningstar database all funds declared S&P 500 as their prospectus benchmark, however, only 460 of those funds fell to the to the Large Cap blend Morningstar category (most aligned to the S&P 500). All the remaining funds were allocated across 21 other distinct Morningstar Global categories, which in terms of fund risk profile and composition are very different from the reported benchmark. Thus, we investigate how significantly the self-reported-benchmark-mismatch affects the inferences on fund performance and whether it can be done for strategic reasons. Preliminary to our analysis we document that Morningstar Global category benchmark (which we call 'true' category benchmark) tested versus the reported S&P 500 benchmark provides a better fit in explaining mutual fund returns (with on average 10% higher R-squared). In addition, similar to Cremers et al. (2012) and Chinthalapati et al. (2017), we report non-zero alphas of passive benchmark indices in our sample, which we adjust as previously with the Angelidis et al. (2013) approach for the unbiased performance evaluation.

To conduct our analysis we split the total sample period into 22 rolling overlapping windows, each being 36 month in length, and estimate the fund performance in relation to S&P 500 versus 'true' category benchmark: first, with the total number of funds in each sub-period (results are obtained with the Carhart, Fama and French three and five-factor models), second, by running individual regressions for each fund in each sub-period (with the Carhart model). Based on 1) similar to Chan et al., (2009) we show that the inferences on mutual fund performance are very sensitive to the benchmarking methodology and model applied. Thus, we document some time periods when the performance of mutual funds benchmarking against S&P 500 instead of the 'true' category benchmark overstates fund performance (for instance in 1997-1999 and 2002-2004, statistically significant at 1% level for all three models). However, there are also cases when the performance assessed against 'true' category benchmark provides better results (e.g. in years 2011-2013, 2012-2014). The results of the individually run regressions for each fund shows that in 70 percent of the periods using S&P 500 as a prospectus benchmark benefits mutual funds with on average higher AGT-adjusted alphas in comparison to the 'true' benchmark. Overall results for the entire periods confirm that Prospectus benchmark overstate the performance of 61.2 % of the funds. Thus, the average AGT-adjusted alpha drops by 23 basis points when 'true' categories benchmark is applied instead.

Furthermore, we estimate the impact of inaccurate benchmark choice on fund relative performance with a particular focus on fund ranking. Our results document that 30% of top performing funds move their ranking position when their performance is adjusted with the 'true' benchmark. Interestingly, our findings reveal that wrong benchmark selection penalises the worst performing funds. Thus, close to 30 % of losers move up their position when performance is estimated with the most suitable global category benchmark. According to outcomes the top quartile funds benefit most from the choice of prospectus benchmark. In some time periods (e.g. 2007-2009) the difference in adjusted alphas in favour of Prospectus benchmarks reaches 460bps. This leads us to conclude that strategic benchmark selection appears to be most likely in the funds at the top performance quartile, while we do not observe clear advantage of benchmark gaming in the remaining quartiles. Overall, the average AGT-alpha when the performance is estimated with 'true' category benchmark drops by 68bps in Quartile 1, falls 28bps and 25bps in Quartiles 2 and 3, and increases 33bps in Quartile 4 in the whole sample period.

In the chapter 5 we analyse mutual fund relative performance and performance persistence in the UK taking into consideration the recent literature highlighting the problems in Fama and French and Carhart portfolio/factor construction and possible mismatches in fund reference benchmarks. To eliminate aforementioned biases in performance assessment we apply the methodology of Hunter, Kandel, Kandel and Wermers (2014) which claims that fund/manager performance should be evaluated in excess of the standard risk factors and commonalities in fund strategies and investment objectives within a peer group. Thereby, the approach proposes to use active peer benchmark (APB) as a passive group reference benchmark, which, according to the methodology, in combination with APB error term represents the fifth factor in addition to the Carhart model and allows to estimate manager skills above the common practices utilised within the group a find belongs. Overall, this method does not aim to estimate the most accurate fund alpha; instead, it enables us to estimate fund relative position in respect to the peers within the reference group. Hence, based on the outcomes we can rank the funds within the reference group and identify subgroups of funds with the top skills.

To conduct the analysis we utilised the sample of 817 active UK long-only equity mutual funds for which we collected net monthly total returns over the period January 1992 to February 2016. Peer-groups were formed based on Morningstar category classification in accordance to fund investment style and size. Following the Hunter et al. (2014) methodology APB return was calculated as the equally weighted return of all the funds in the same Morningstar category, therefore, this method elominates potential biases in fund relative assessment caused by mismatched self-reported benchmarks. Based on the APB adjusted Carhart model we obtained APB adjusted alphas for each mutual fund and split the results into four quartiles (by t-test) in order to identify the top and the bottom performing funds. In addition, we performed test for persistence in performance of the most skilled and the bottom quartile funds one-year-ahead using funds' excess returns and Carhart alphas.

The results obtained from both parametric (regression) and non-parametric (contingency tables) methods showed the evidence of performance persistence regardless of the method employed. The outcomes remained robust even when we split the funds into deciles. In contrast to the exiting literature we document the performance persistence for both: winner and loser funds, with the result being consistent across all Morningstar peer groups. Summarising the results we posit that APB methodology/ APB adjusted alphas have strong predictive ability of future returns. In support of this statement we provide evidence that 1% increase in t-statistics of APB-adjusted alphas leads to 2.37% increase in probability that a fund will be placed in the top performing quartile.

Overall, in this dissertation we emphasise that in order to conduct unbiased mutual fund performance evaluation the outcomes obtained with the standard Fama and French and Carhart models should be adjusted for non-zero benchmark alphas and should be measured relative to the most accurate reference benchmark. Thus, the previous empirical studies based on standard performance measures strongly tilt towards significant underperformance of UK funds. In fact, we provide evidence that after the passive benchmark alpha adjustment UK focused equity funds are able to outperform even during bear market periods. Moreover, the fund performance persists for the winner and loser funds one-year ahead.

Benchmarking against unsuitable benchmark may significantly overstate the performance of winners and underestimate the performance of losers. The fact that the fund self-declared prospectus benchmark does not represent funds' actual investment style in 2/3 of the funds in our sample should bring attention of financial regulators and policy makers. The new information disclosure requirements need to be put in place in order to provide more clarity for investors on how the prospectus benchmark is selected. Until then, investors have to be more cautious when interpreting performance figures in fund prospectus. As an alternative, APB methodology can be used to estimate a fund performance within a peer group, which allows ranking the funds by manager skills. As a solution to self-reported-benchmark-mismatch problem, policy makers could request that the fund identifies its peer-group and that performance relative to that peer group is made available to investors.

Bibliography

Aharoni, G., Grundy, B. and Zeng, Q., 2013. Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. Journal of Financial Economics, 110(2), pp.347-357.

Ang, A., R. J. Hodrick, J. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. Journal of Finance 51: 259–299.

Angelidis T., Giamouridis D., Tessaromatis N., 2013. Revisiting mutual fund performance evaluation Journal of Banking & Finance Volume 37, Issue 5, May 2013, Pages 1759–1776

Arshanapalli, Bala, T. Daniel Coggin, and John Doukas, 1998. Multifactor asset pricing analysis of international value investment strategies, The Journal of Portfolio Management 24, 10-23.

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. The Journal of Finance Volume 68, Issue 3, pages 929–985.

Ball, R., 1978. "Anomalies in Relationships Between Securities' Yields and Yieldsurrogates", Journal of Financial Economics, 6, no. 2/3: pp 103-126

Ball, R., Gerakos, J., Linnainmaa, J.T. and Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. Journal of Financial Economics, 121(1), pp.28-45.

Bams, D., Otten, R. and Ramezanifar, E., 2016. Investment Style Misclassification and Mutual Fund Performance, working paper.

Banerjee P. S., Doran J. S., Peterson D. R., 2007. Implied volatility and future portfolio returns Journal of Banking & Finance, Volume 31, Issue 10, October, Pages 3183–3199

Banz, R. W., 1981. “The relationship between return and market value of common stocks”, Journal of Financial Economics, Vol. 9, Issue 1, pp. 3-18

Barras L.,Scaillet O. and Wermers R., 2010. False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas The Journal of Finance, Volume 65, Issue 1, pages 179–216, February 2010

Bartholdy, J., Peare, P., 2005. Estimation of expected return: CAPM versus fama and french. Int. Rev. Financial Anal. 14 (4), 407–427.

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

Ben-Rephael, A., Kandel, S. and Wohl, A., 2012. Measuring investor sentiment with mutual fund flows. Journal of Financial Economics, 104(2), pp.363-382.

Berk, J., R. Green, and V. Naik. 1999. Optimal investment, growth options, and security returns. Journal of Finance 54 (October): 1553-1607.

Bhandari, L. C., 1988. "Debt/equity ratios and expected common stock returns: Empirical evidence" Journal of Finance 43, 507-528

Blake, D., Caulfield, T., Ioannidis, C. and Tonks, I., 2014. Improved inference in the evaluation of mutual fund performance using panel bootstrap methods. Journal of econometrics, 183(2), pp.202-210.

Blake, D. and Timmerman, A., 1998. Mutual fund performance: evidence from the UK. European Finance Review 2, pp. 57–77.

Bollen, N., Busse, J.A., 2005. Short-term persistence in mutual fund performance. Review of Financial Studies 18, 569–597.

Bourguignon F., Jong M., 2003. "Value versus growth" The Journal of Portfolio Management, Vol. 29, No. 4: pp. 71-79

Brennan, Michael, Ashley Wang, and Yihong Xia, 2004. Estimation and test of a simple model of intertemporal asset pricing. Journal of Finance 59, 1743-1775.

Brown S. J. and Goetzmann W. N., 1995. Performance Persistence, Journal of Finance, 50 (2), pp. 679-698.

Brown, S.J. and Goetzmann, W.N., 1997. Mutual fund styles. Journal of financial Economics, 43(3), pp.373-399.

Brown, Keith C., W. V. Harlow, and Laura T. Starks, 1996. Of tournaments and temptations: An analysis of managerial incentives in mutual fund industry, Journal of Finance 51, 85–110.

Byrne A., Fletcher J and Ntozi P., 2006. An Exploration of the Conditional Timing Performance of UK Unit Trusts Journal of Business Finance & Accounting Volume 33, Issue 5-6, pages 816–838, June/July 2006

Carhart, M. M., 1997. On persistence in mutual fund performance, Journal of Finance 52:57-82.

Castellanos, A.R. and Alonso, B.V., 2005. Spanish Mutual Fund Misclassification: Empirical Evidence. The Journal of Investing, 14(1), pp.41-51.

Chalmers J. M. R. , Edelen R. M., and Kadlec G. B., 2001. Fund Returns and Trading Expenses: Evidence on the Value of Active Fund Management, Working Paper

Campbell, J.Y., Hilscher, J. and Szilagyi, J., 2008. In search of distress risk. The Journal of Finance, 63(6), pp.2899-2939.

Chan L. K. C. and Lakonishok J., 2004. Value and Growth Investing: Review and Update, Financial Analysts Journal, 60(1), pp.71-86

Chan L. K. C., Chen H., and Lakonishok J., 2002. On Mutual Fund Investment Styles, Rev. Financ. Stud. (2002) 15 (5):1407-1437.

Chan L.K.C. , Dimmock S. G. and Lakonishok J., 2009. Benchmarking Money Manager Performance: Issues and Evidence The Review of Financial Studies,22 (11):4553-4599.

Chan, Louis K.C., Yasushi Hamao, and Josef Lakonishok, 1991. Fundamentals and stock returns in Japan, Journal of Finance 46, 1739-1789

Chen, J., H. Hong, M. Huang and J. D. Kubik, 2004. "Does fund size erode mutual fund performance? The role of liquidity and organization", American Economic Review, 94(5): 1276-1302

Chen, Z. and Knez P.J., 1996. Portfolio Performance Measurement: Theory and Applications, Review of Financial Studies, 2, pp. 511-555.

Chen, L., Novy-Marx, R. and Zhang, L., 2011. An alternative three-factor model. Working paper

Chevalier J.and Ellison G., 1997. Risk taking by mutual funds as a response to incentives**,** Journal of Political Economy, 105, pp. 1167–1200

Chiah, M., Chai, D., Zhong, A., Li, S., 2016. A better model? An empirical investigation of the Fama-French five-factor model in Australia. Int. Rev. Financ. 16 (4)

Chinthalapati, V.L., Mateus, C. and Todorovic, N., 2017. Alphas in disguise: A new approach to uncovering them, International Journal of Finance and Economics, 22 (3), pp. 234-243.

Costa, B. and K. Jakob, 2006. Do Stock indexes Have Abnormal Performance?, The Journal of Performance Measurement 11 (1): 8-18

Cremers, K.M. and Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies, 22(9), pp.3329-3365.

Cremers, M., Petajisto, A. and Zitzewitz, E., 2012. Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, Critical Finance Review, 2, pp.1-48

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N., 2010. Mutual fund performance:Measurement and evidence. Journal of Financial Markets, Instruments and Institutions, 19(2), 95–187.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N., 2012. False discoveries in UK mutual fund performance. European Financial Management, 18(3), pp. 444–463

Cuthbertson, K., Nitzsche, D., O'Sullivan, N., 2008. UK mutual fund performance: skill or luck? Journal of Empirical Finance, 15 (4), pp. 613–634.

Dan diBartolomeo and Witkowski, E., 1997. Mutual fund misclassification: Evidence based on style analysis. Financial Analysts Journal, pp.32-43.

Daniel, K. and Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. the Journal of Finance, 52(1), pp.1-33.

Daniel, K., Grinblatt M., Titman S. and Wermers R., 1997. "Measuring Mutual Fund Performance with Characteristics-Based Benchmarks", Journal of Finance, Vol. 52, pp.1035–1058

Dimpson E, Marsh P. and M. Staunton, 2004. Low-Cap and Low-Rated Companies: Tiny and Troubled in 20th Century Europe, Journal of Portfolio Management, 30 (4), pp.133-143.

Dimson E. and Marsh P., 2001. U.K. Financial Market Returns, 1955–2000, The Journal of Business, Vol. 74, No. 1, pp. 1-31

Drew, M. E., and Veeraraghavan, M., 2002. "A closer look at the size and value premium in emerging markets: Evidence from the Kuala Lumpur Stock Exchange", Asian Economic Journal, Vol. 16, pp.337–351

Eckbo B. E. and Norli Ø., 2002. Pervasive liquidity risk Working paper, Tuck School of Business at Dattmouth

Eckbo B. E. and Norli Ø., 2005. Liquidity risk, leverage and long-run IPO returns. Journal of Corporate Finance, Volume 11, Issues 1–2, March 2005, Pages 1–35

Elton, E., Gruber, M., Blake, C., 2003. Incentive fees and mutual funds. Journal of Finance 58, pp.779–804.

Faff, R. 2001. "An Examination of the Fama and French three-factor model using commercially available factors", Australian Journal of Management, Vol. 26, pp.1-17

Fairfield, P.M., Whisenant, J.S. and Yohn, T.L., 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. The accounting review, 78(1), pp.353-371.

Fama E.F. and French K.R., 1993. "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics, Vol. 33, pp.3-56

Fama E.F. and French K.R., 1996. "Multifactor Explanations of Asset-Pricing Anomalies", Journal of Finance, Vol. 47, pp.426-465

Fama, E.F. and French, K.R., 2008. Dissecting anomalies. The Journal of Finance, 63(4), pp.1653-1678.

Fama, E.F. and K.R. French, 2012. `Size, Value, and Momentum in International Stock Returns‟,Journal of Financial Economics Volume 105, Issue 3, September, Pages 457–472

Fama, Eugene and Macbeth, J., 1973. "Risk, Return and Equilibrium: Empirical Tests" The Journal of Political Economy Vol. 81, No 3 (May-Jun, 1973), 607-636

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, Journal of Finance 47, 427-465.

Fama, E. , French, K. , 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22 .

Fama, E. , French, K. , 2016. Dissecting anomalies with a five-factor model. Review of Financial Studies 29, 70–103 .

Fama, E. , French, K. , 2017 International tests of a five-factor asset pricing model Journal of Financial Economics Volume 123, Issue 3, 441–463

FCA Asset Management Market Study, Final Report (June, 2017): last accessed (26.08. 2017) at: https://www.fca.org.uk/publication/market-studies/ms15-2-3.pdf

Ferguson, Michael F., and Richard L. Shockley, 2003. Equilibrium anomalies, Journal of Finance 58, 2549–2580.

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

Fletcher, J. and D. Forbes, 2002. "An Exploration of the Persistence of UK Unit Trusts Performance. Journal of Empirical Finance, 9, pp. 475-493.

Foran, J. and O'Sullivan, N., 2014. Liquidity risk and the performance of UK mutual funds. International Review of Financial Analysis, 35, pp.178-189.

French, Kenneth R., 2008. The Cost of Active Investing, Journal of Finance 63, 1537–1573.

Frijns, B., Gilbert, A. and Zwinkels, R.C., 2016. On the Style-Based Feedback Trading of Mutual Fund Managers. Journal of Financial and Quantitative Analysis, 51(3), pp.771-800.

Fung, W. and Hsieh, D.A., 2002. Hedge-fund benchmarks: Information content and biases. Financial Analysts Journal, pp.22-34.

Gharghori P. ,Chan H., Faff R., 2007. Are the Fama-French Factors Proxying Default Risk? Australian Journal of Management December, vol. 32 no. 2 223-249

Goetzmann, W., Ingersoll, J., Spiegel, M. and Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies, 20(5), pp.1503-1546.

Gregory A, Tharayan R. and A. Christides, 2013. Constructing and Testing Alternative Versions of the Fama–French and Carhart Models in the UK, Journal of Business Finance & Accounting, 40: 172–214

Griffin, J. M., and M. Lemmon, 2002."Book–to–Market Equity, Distress Risk, and Stock Returns," The Journal of Finance Volume 57, Issue 5, pages 2317–2336

Griffin, J.M. 2002. 'Are the Fama and French Factors Global or Country Specific?' Review of Financial Studies, Vol. 15, No. 3, pp. 783–803.

Grinblatt M., Titman S. and Wermers R., 1995. Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, The American Economic Review, 85(5): 1088-1105

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

Herrmann U., Scholz H., 2013. Short-term persistence in hybrid mutual fund performance: The role of style-shifting abilities Journal of Banking & Finance, Volume 37, Issue 7, pp. 2314–2328

Hou K., Karolyi G.A., Kho B.C., 2011. What factors drive global stock returns? Review of Financial Studies, 24, pp.2527-2574.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28 (3), 650–705.

Huynh, T.D., 2017. Explaining Anomalies in Australia with a Five-factor Asset Pricing Model. International Review of Finance. Forthcoming

Hsu J.C., Kalesnik V., and Myers B. W., 2010. Performance Attribution: Measuring Dynamic Allocation Skill Financial Analysts Journal, Volume 66 • Number 6

Huang, J., Sialm, C. and Zhang, H., 2011. Risk shifting and mutual fund performance. Review of Financial Studies, 24(8), pp.2575-2616.

Huij, J.J. & Verbeek, M.J.C.M., 2009. On the use of multifactor models to evaluate mutual fund performance, Financial Management, 38 (1): 75-102

Hunter, D., Kandel E., Kandel S. and Wermers R., 2014. Mutual fund performance evaluation with active peer benchmarks, Journal of Financial Economics, 112(1), pp. 1-29.

Ikenberry, D., Lakonishok, J. and Vermaelen, T., 1995. Market underreaction to open market share repurchases. Journal of financial economics, 39(2), pp.181-208.

James L. Davis J. L., 2001. Mutual Fund Performance and Manager Style Financial Analysts Journal, Vol. 57, No. 1 (Jan. - Feb.), pp. 19-27

Jegadeesh N. and Titman S., 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", The Journal of Finance, Vol. 48, pp. 65-91

Jennifer H., Sialm C., and Zhang H., 2011. Risk shifting and mutual fund performance, Review of Financial Studies 24, 2575–2616.

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

Jiang W., 2003. A nonparametric test of market timing, Journal of Empirical Finance**,** Volume 10, Issue 4, September 2003, Pages 399–425

Kacperczyk M., Sialm C. and Zheng L., 2005. On the Industry Concentration of Actively Managed Equity Mutual Funds, The Journal of Finance Volume 60, Issue 4, pages 1983–2011, August 2005

Kacperczyk, M., Sialm, C. and Zheng, L., 2008. Unobserved actions of mutual funds. The Review of Financial Studies, 21(6), pp.2379-2416.

Kacperczyk M.,Nieuwerburgh S. V. and Veldkamp L., 2014. Time-Varying Fund Manager Skill The Journal of Finance, Volume 69, Issue 4, pages 1455–1484.

Karoui A. and Meier I., 2009, Performance and characteristics of mutual fund starts, The European Journal of Finance,Volume 15, Issue 5-6

Keim, D.B., 1983. “Size-related Anomalies and Stock Return Seasonality: Further Empirical Evidence.” Journal of Financial Economics 12(1): 13 – 32.

Kim, Dongcheol, and Myung-sun Kim, 2003. A multi-factor explanation for postearnings-announcement drift. Journal of Financial and Quantitative Analysis 38, 383- 398.

Kim, M., Shukla, R. and Tomas, M., 2000. Mutual fund objective misclassification. Journal of Economics and Business, 52(4), pp.309-323.

Kosowski, R., A. Timmermann, H. White and R. Wermers. 2006. “Can Mutual Fund “Stars” Really Pick Stocks? New Evidence from a Bootstrap Analysis.” Journal of Finance LX1:6:2551-2595.

Kosowski, Robert, 2011. Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions, Quarterly Journal of Finance 1, 607–664.

Kothari, S.P. and Warner, J.B., 1997. Measuring long-horizon security price performance. Journal of Financial Economics, 43(3), pp.301-339.

L'Her, J., Masmoudi T. and Suret J., 2004. "Evidence to Support the Four-Factor Pricing Model from the Canadian Stock Market", Journal of International Financial Markets, Vol. 14, pp.313–328

Lam, K.S.K, Li, F.K. and So, S.M.S., 2010. "On the validity of the augmented Fama and French's (1993) model: evidence from the Hong Kong stock market", Review of Quantitative Finance and Accounting, Vol. 35, pp.89-111

Levis M. and Liodakis M., 1999. The Profitability of Style Rotation Strategies in the United Kingdom, The Journal of Portfolio Management Fall 1999, Vol. 26, No. 1: pp. 73-86

Liew, J. and Vassalou M., 2000. "Can Book-to-Market, Size and Momentum be Risk Factors that Predict Economic Growth?", Journal of Financial Economics, Vol. 57(2), pp.221–245

Lin, Q., 2017. Noisy prices and the Fama–French five-factor asset pricing model in China. Emerging Markets Review, 31, pp.141-163

Lintner J., 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets" Review of Economics and Statistics, Vol. 47, pp. 13–37

Liu, W., 2006. "A Liquidity-Augmented Capital Asset Pricing Model." Journal of Financial Economics, vol. 82, no. 3 (December): 631-671.

Loughran, T. and Ritter, J.R., 1995. The new issues puzzle. The Journal of finance, 50(1), pp.23-51.

Lozano B., 2006. "Estimating and evaluating the Fama-French & Carhart models", working paper

Lucas, A., van Dijk, R., & Kloek, T., 2002. Stock selection, style rotation, and risk. Journal of Empirical Finance, 9(1), 1–34.

Mamaysky, H., Spiegel, M. and Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. The Review of Financial Studies, 21(1), pp.233-264.

Markowitz, H. 1952 "Portfolio Selection" The Journal of Finance, Vol. 7, No. 1,pp. 77-91

Maroney, N. and Protopapadakis A., 2002. "The Book-to-Market and Size Effects in a General Asset Pricing Model: Evidence from Seven National Markets", European Finance Review, Vol. 6, pp.189-221

Monney, A., 2016. Passive funds grow 230% to $6 tn. Financial Times, last accessed (26.08. 2017) at: https://www. ft. com/content/2552ce62-2400-11e6-aa98-db1e01fabc0c.

Moreno D. , Rodríguez R., 2009. The value of coskewness in mutual fund performance evaluation Journal of Banking & Finance, September, Pages 1664–1676

Moskowitz, T, 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion, Journal of Finance 55, 1695–1703.

Nichol, E. and Dowling, M., 2014. Profitability and investment factors for UK asset pricing models. Economics Letters, 125(3), pp.364-366.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1–28.

Otten, R., & Reijnders, M., 2012. The performance of small cap mutual funds: Evidence for the UK. Working paper Maastricht University Department of Finance

Pastor, L. and R. F. Stambaugh, 2002. Mutual fund performance and seemingly unrelated assets, Journal of Financial Economics 63: 315-349.

Pastor, L., Stambaugh,R.F., 2003. Liquidity risk and expected stock returns. J.Polit.Econ.111(3),642–685.

Petajisto A., 2013. Active Share and Mutual Fund Performance Financial Analysts Journal Volume 69 Number 4, pp. 73-93

Petersen, M. A., 2009. "Estimating standard errors in finance panel data sets: Comparing approaches", Review of Financial Studies, 22: 435-480

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

Reinganum, M. R., 1999, "The Significance of Market Capitalisation in Portfolio Management Over Time", Journal of Portfolio Management, Summer 1999, pp. 39-50.

Sadka, Ronnie, 2006. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, Journal of Financial Economics 80, 309–349

Sensoy, B.A., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry. Journal of Financial Economics, 92(1), pp.25-39.

Sharpe W. F., 1991. The Arithmetic of Active Management Financial Analysts Journal Vol. 47, No. 1 (Jan. - Feb., 1991), pp. 7-9

Sharpe, W. F., 1964. “Capital asset prices: A theory of market equilibrium under conditions of risk” Journal of Finance 19, 425-42

Shukla, R., 2004. “The Value of Active Portfolio Management. Journal of Economics and Business 56:331-346.

Simlai, P.E., 2016. Time-varying risk, mispricing attributes, and the accrual premium. International Review of Financial Analysis, 48, pp.150-161.

Sloan, R.G., 1996. “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?” The Accounting Review 71: 289-315.

Spiess, D.K. and Affleck-Graves, J., 1995. Underperformance in long-run stock returns following seasoned equity offerings. Journal of Financial Economics, 38(3), pp.243-267.

Swinkels, L. and Van Der Sluis, P.J., 2006. Return-based style analysis with time-varying exposures. The European Journal of Finance, 12(6-7), pp.529-552.

Vassalou, M., and Y. Xing, 2004, “Default risk in equity returns,” Journal of Finance, 59, 831–868.

Zaremba, A. and Czapkiewicz, A., 2017. Digesting anomalies in emerging European markets: A comparison of factor pricing models. Emerging Markets Review, 31, pp.1-15.

Wermers R., 2000. Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses, The Journal of Finance, 55 (4): 1655–1703.

Wermers, R., 2012. Matter of style: The causes and consequences of style drift in institutional portfolios. Working paper.

Zhang, C., 2008. Decomposed Fama-French factors for the size and book-to-market effects. Working paper, Hong Kong University of Science and Technology. Working paper

Zheng, L., 1999. Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. Journal of Finance 54:901-33.

Appendix:

Table A1: (Chapter 4) Difference in alphas per quartile and change of quartile ranks estimated with FF3 and FF5 factor models

Panels A-H report results for Quartile 1(top) - 4 (bottom) respectively. All panels show the number of funds and comparison of AGT adjusted alphas, when S&P 500 is used as a benchmark ($\alpha_i^{*S\&P500}$ from eq(5)) and when 'true' benchmark is used (α_i^{*True} from equation (6)). Alphas and the difference in alphas are annualised and given in basis points. The last column shows percentage of funds that remains in the same quartile when the benchmark is changed from the S&P500 to the 'true' benchmark. In the last row, the 'average' represents the average across the periods and across the funds.

Panel A: Quartile 1 (FF3 factor model)

Years	*# Funds*	*Average Annual Alpha S&P500*	*Average Annual Alpha Global Category*	*difference*	*% Funds same Quartile*
199201:199412	48	819.6278	686.2782	-133.35	83.33
199301:199512	55	922.4035	691.9723	-230.431	74.55
199401:199612	61	480.9575	534.8321	53.8746	60.66
199501:199712	69	443.8568	538.9791	95.1223	68.12
199601:199812	75	534.3349	529.5584	-4.7765	44.00
199701:199912	86	866.613	717.2576	-149.355	75.58
199801:200012	96	1985.415	1588.041	-397.374	84.38
199901:200112	108	1548.902	1356.564	-192.338	82.41
200001:200212	119	1051.343	1212.398	161.055	71.43
200101:200312	132	738.6076	934.7219	196.1143	73.48
200201:200412	134	476.0666	520.0906	44.024	66.42
200301:200512	129	647.6853	430.0229	-217.662	74.42
200401:200612	131	573.5268	490.8361	-82.6907	74.81
200501:200712	129	1297.91	676.4457	-621.464	64.34
200601:200812	128	886.9872	477.587	-409.4	68.75
200701:200912	128	1081.405	555.8459	-525.559	67.19
200801:201012	127	976.2867	609.8394	-366.447	75.59
200901:201112	123	727.8809	601.6418	-126.239	56.91
201001:201212	122	401.5851	492.1938	90.6087	68.85
201101:201312	121	420.1622	523.0454	102.8832	61.98
201201:201412	118	465.1062	-284.099	-749.205	22.88
201301:201602	112	252.2624	221.9613	-30.3011	66.96

Panel B: Quartile 2 (FF3 factor model)

Years	*# Funds*	*Average Annual Alpha S&P500*	*Average Annual Alpha Global Category*	*difference*	*% Funds same Quartile*
199201:199412	48	88.15998	100.6615	12.50152	50.00
199301:199512	54	71.52094	90.85617	19.33523	51.85
199401:199612	61	-74.5937	7.670695	82.2644	32.79
199501:199712	69	-28.1794	71.0302	99.2096	40.58
199601:199812	75	-9.53839	-68.186	-58.6476	5.33
199701:199912	86	-82.9789	-111.031	-28.0521	34.88
199801:200012	96	419.523	320.9465	-98.5765	68.75
199901:200112	108	361.5625	322.0583	-39.5042	62.96
200001:200212	119	245.8971	122.1182	-123.779	71.43
200101:200312	131	-97.6515	-89.6297	8.0218	70.99
200201:200412	133	-22.4605	-19.1831	3.2774	45.86
200301:200512	128	112.927	-11.9934	-124.92	53.13
200401:200612	131	129.1343	36.05946	-93.0748	61.83
200501:200712	128	409.3904	183.9326	-225.458	61.72
200601:200812	128	184.5661	88.19451	-96.3716	71.88
200701:200912	128	262.4373	98.48174	-163.956	47.66
200801:201012	126	115.1935	57.14915	-58.0444	70.63
200901:201112	122	97.17655	-2.87962	-100.056	43.44
201001:201212	122	-50.9551	-32.3519	18.6032	63.93
201101:201312	121	-68.1577	8.811557	76.96926	61.98
201201:201412	118	-50.3335	32.08709	82.42059	30.51
201301:201602	112	-130.212	-126.461	3.751	62.50

Panel C: Quartile 3 (FF3 factor model)

Years	*# Funds*	*Average Annual Alpha S&P500*	*Average Annual Alpha Global Category*	*difference*	*% Funds same Quartile*
199201:199412	47	-127.089	-121.556	5.533	44.68
199301:199512	54	-179.897	-141.3	38.597	51.85
199401:199612	62	-301.777	-260.537	41.24	35.48
199501:199712	68	-311.592	-242.986	68.606	48.53
199601:199812	74	-335.839	-367.379	-31.54	14.86
199701:199912	86	-473.688	-418.649	55.039	32.56
199801:200012	96	-55.0246	-86.369	-31.3444	63.54
199901:200112	109	-31.8455	-39.162	-7.3165	76.15
200001:200212	119	-139.184	-299.536	-160.352	73.95
200101:200312	131	-466.502	-471.278	-4.776	68.70
200201:200412	133	-261.843	-286.107	-24.264	51.88
200301:200512	128	-109.025	-235.838	-126.813	50.78
200401:200612	131	-78.7528	-164.51	-85.7572	66.41
200501:200712	128	97.54926	-50.8165	-148.366	53.91
200601:200812	129	-52.5598	-118.696	-66.1362	73.64
200701:200912	129	45.03163	-123.001	-168.033	48.06
200801:201012	126	-115.23	-158.483	-43.253	76.19
200901:201112	122	-161.01	-275.734	-114.724	53.28
201001:201212	121	-296.962	-265.732	31.23	65.29
201101:201312	121	-367.441	-240.582	126.859	57.02
201201:201412	119	-278.504	-230.72	47.784	30.25
201301:201602	113	-339.274	-300.967	38.307	63.72

Panel D: Quartile 4 (FF3 factor model)

Years	*# Funds*	*Average Annual Alpha S&P500*	*Average Annual Alpha Global Category*	*difference*	*% Funds same Quartile*
199201:199412	48	-565.864	-612.647	-46.783	75.00
199301:199512	55	-646.859	-648.459	-1.6	61.82
199401:199612	61	-892.085	-840.535	51.55	59.02
199501:199712	69	-1434.2	-1159.56	274.64	66.67
199601:199812	75	-1211.57	-1074.47	137.1	58.67
199701:199912	86	-1454.2	-1515.43	-61.23	74.42
199801:200012	96	-805.213	-832.482	-27.269	75.00
199901:200112	108	-646.261	-741.973	-95.712	76.85
200001:200212	119	-881.168	-980.28	-99.112	73.11
200101:200312	132	-1070.61	-1141.4	-70.79	79.55
200201:200412	134	-771.368	-798.096	-26.728	79.10
200301:200512	129	-504.003	-600.844	-96.841	66.67
200401:200612	131	-563.016	-670.566	-107.55	80.15
200501:200712	129	-405.685	-466.987	-61.302	61.24
200601:200812	128	-517.529	-492.042	25.487	78.91
200701:200912	128	-420.235	-494.435	-74.2	64.06
200801:201012	127	-569.725	-602.543	-32.818	80.31
200901:201112	123	-657.379	-754.692	-97.313	69.11
201001:201212	122	-912.704	-862.462	50.242	77.05
201101:201312	121	-1349.32	-860.72	488.6	61.16
201201:201412	118	-1308.82	-244.886	1063.934	31.36
201301:201602	112	-1059.6	-700.358	359.242	71.43

Panel E: Quartile 1 (Fama and French Five Factor Model)

Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} - $\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 1*
199201:199412	48	1234.89	889.4794	-345.41	83.33
199301:199512	55	1248.249	890.4617	-357.787	70.91
199401:199612	61	757.193	657.5208	-99.6722	73.77
199501:199712	69	1025.816	864.774	-161.042	75.36
199601:199812	75	953.4566	809.7909	-143.666	58.67
199701:199912	86	881.1615	747.5386	-133.623	63.95
199801:200012	96	1867.395	1485.052	-382.343	78.13
199901:200112	108	1543.719	1437.234	-106.484	76.85
200001:200212	119	1638.89	1562.894	-75.9959	73.11
200101:200312	132	1126.569	1001.424	-125.145	72.73
200201:200412	134	756.2595	605.0557	-151.204	74.63
200301:200512	129	728.6072	488.088	-240.519	79.07
200401:200612	131	593.4402	451.5606	-141.88	77.86
200501:200712	129	1037.853	570.0174	-467.836	68.22
200601:200812	128	1094.703	540.2681	-554.435	66.41
200701:200912	128	1283.563	745.6192	-537.944	63.28
200801:201012	127	1338.104	843.9821	-494.122	80.31
200901:201112	123	951.4045	690.0758	-261.329	65.04
201001:201212	122	637.8727	454.6618	-183.211	72.95
201101:201312	121	465.6073	427.1907	-38.4165	84.30
201201:201412	118	376.2405	348.7859	-27.4546	79.66
201301:201602	112	312.8956	282.5858	-30.3098	77.68
			Average	**-299.99**	**73.46**

Panel F: Quartile 2 (Fama and French Five Factor Model)

Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} - $\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 2*
199201:199412	48	312.3043	224.4458	-87.8585	60.42
199301:199512	54	252.6547	133.4499	-119.205	57.41
199401:199612	61	28.97568	-6.57166	-35.5473	55.74
199501:199712	69	242.6822	166.7283	-75.9539	66.67
199601:199812	75	163.3163	43.54876	-119.768	17.33
199701:199912	86	-106.654	-98.2076	8.446284	30.23
199801:200012	96	311.2834	290.6146	-20.6688	50.00
199901:200112	108	243.0482	202.5278	-40.5204	58.33
200001:200212	119	461.97	149.9872	-311.983	50.42
200101:200312	131	-63.0407	-80.1895	-17.1488	44.27
200201:200412	133	64.27896	-43.7514	-108.03	57.89
200301:200512	128	152.1308	-32.0181	-184.149	58.59
200401:200612	131	117.7403	-1.31765	-119.058	61.83
200501:200712	128	339.0442	134.0216	-205.023	54.69
200601:200812	128	291.5599	118.7052	-172.855	58.59
200701:200912	128	383.4283	157.8579	-225.57	35.94
200801:201012	126	309.3071	174.8805	-134.427	57.14
200901:201112	122	199.4383	70.19669	-129.242	46.72
201001:201212	122	102.3873	57.72158	-44.6658	54.10
201101:201312	121	24.68945	24.47427	-0.21518	74.38
201201:201412	118	-33.8305	-54.0718	-20.2413	77.97
201301:201602	112	-126.685	-113.218	13.46698	73.21
			Average	**-97.74**	**54.63**

Panel G: Quartile 3 (Fama and French Five Factor Model)

Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} -$\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 3*
199201:199412	47	-18.303	-36.5895	-18.2865	63.83
199301:199512	54	-83.8714	-128.096	-44.2248	70.37
199401:199612	62	-237.079	-270.491	-33.4113	58.06
199501:199712	68	-90.7987	-161.024	-70.2255	67.65
199601:199812	74	-186.177	-299.317	-113.14	10.81
199701:199912	86	-525.195	-452.838	72.35717	25.58
199801:200012	96	-223.85	-181.769	42.08035	41.67
199901:200112	109	-208.923	-214.207	-5.28374	62.39
200001:200212	119	-38.6113	-348.37	-309.759	47.90
200101:200312	131	-374.405	-484.811	-110.405	48.85
200201:200412	133	-203.585	-272.956	-69.3712	56.39
200301:200512	128	-78.299	-254.383	-176.084	60.94
200401:200612	131	-100.847	-215.6	-114.753	63.36
200501:200712	128	71.98166	-67.3231	-139.305	53.13
200601:200812	129	38.11069	-79.2032	-117.314	67.44
200701:200912	129	93.15661	-120.298	-213.455	43.41
200801:201012	126	12.86472	-105.683	-118.548	54.76
200901:201112	122	-75.8051	-207.569	-131.764	54.10
201001:201212	121	-108.276	-156.133	-47.8565	47.93
201101:201312	121	-209.699	-190.167	19.53125	75.21
201201:201412	119	-246.686	-254.489	-7.8031	76.47
201301:201602	113	-330.183	-299.326	30.85763	76.11
			Average	**-76.19**	**55.74**

Panel H: Quartile 4 (Fama and French Five Factor Model)

Period	*# of Funds*	*Average* $\alpha_i^{*S\&P500}$ *(bp)*	*Average* α_i^{*True} *(bp)*	*Average difference* α_i^{*True} - $\alpha_i^{*S\&P500}$ *(bp)*	*% Funds remaining in Quartile 4*
199201:199412	48	-448.988	-674.337	-225.349	81.25
199301:199512	55	-624.371	-682.079	-57.7084	74.55
199401:199612	61	-1017.04	-961.176	55.86681	78.69
199501:199712	69	-1026.96	-1013.93	13.028	79.71
199601:199812	75	-854.679	-928.069	-73.3907	44.00
199701:199912	86	-1554	-1520.24	33.76355	65.12
199801:200012	96	-999.041	-1132.02	-132.978	64.58
199901:200112	108	-856.667	-1013.93	-157.264	70.37
200001:200212	119	-758.82	-1064.65	-305.828	69.75
200101:200312	132	-911.208	-1084.27	-173.063	68.94
200201:200412	134	-577.366	-689.449	-112.084	76.87
200301:200512	129	-440.964	-671.095	-230.13	69.77
200401:200612	131	-676.381	-752.039	-75.6578	74.05
200501:200712	129	-389.113	-460.467	-71.3539	69.77
200601:200812	128	-463.463	-465.994	-2.53099	71.09
200701:200912	128	-413.498	-527.48	-113.982	56.25
200801:201012	127	-429.051	-602.963	-173.912	77.95
200901:201112	123	-573.091	-711.584	-138.493	77.24
201001:201212	122	-595.056	-727.546	-132.489	72.95
201101:201312	121	-1051.81	-740.747	311.0593	74.38
201201:201412	118	-1143.74	-771.588	372.1481	75.42
201301:201602	112	-1077.56	-719.625	357.9346	75.89
			Average	**-46.93**	**71.30**