Review of new trends in the literature on factor models and mutual fund performance

Irina B. Mateus[a],

Cesario Mateus[b],

Natasa Todorovic[c*],

[a] *Unaffiliated*

[b] *University of Roehampton Business School*

[c] *Cass Business School, City, University of London*

Abstract

In this paper we provide critical review of recent developments in the mutual fund performance evaluation literature. The new literature centres around two main themes: enhancing explanatory power of the standard Fama-French-Carhart factor models by augmenting them with different factors and altering standard models to account for presence of non-zero alphas in passive indices used as fund benchmarks. The latter includes the literature providing solutions for scenarios in which those benchmarks do not match fund objectives. We find that in the plethora of suggested 'missing' factors, not one can be universally used to explain all anomalies or price all stocks. We also find that new models that adjust a fund's standard Carhart alpha for alpha of its benchmark or for commonalities in its peer–group, provide additional information on fund performance to that given by the standard models. Specifically, these models give account of fund's *relative* performance - to the benchmark or the peer-group, which is of use to investors.

Keywords: Standard factor models, Mutual fund performance, Augmented models, Benchmark-adjusted models, Peer-group adjusted models

JEL classification: G1, G23

---

*Corresponding author. Tel.: +44 207 040 0120. Address: 106 Bunhill Row, London EC1Y 8TZ, UK. *E-mail addresses:* bzhirina@gmail.com (I. B. Mateus), cesario.mateus@roehampton.ac.uk (C. Mateus), n.todorovic@city.ac.uk (N. Todorovic).

Contents

1. Introduction

With global asset and wealth management industry predicted to grow exponentially from $84.9trn in 2016 to $145.4trn in 2025[1], and over 9300 mutual funds in the US alone[2], the need for unbiased performance evaluation becomes increasingly important. A significant strand of past academic literature has looked at the performance of actively managed equity funds, deploying the standard Fama-French and Carhart factor models. The prevailing results point at funds' underperformance, i.e. a negative after-fee alpha (see for instance Gruber 1996, Carhart, 1997, Fama and French 2010, Berk and van Binsbergen 2012 etc.). Only a small number of (statistically and economically) significant mutual funds generate positive alphas (e.g. Wermers (2000), Kosowsky et al., 2006), and once false discoveries are taken into account, no more than 5% of funds can be considered as true winners (Cuthbertson et al., 2010). Further, the studies are by and large in agreement that performance of winner funds does not persist, signalling to investors that ex-ante selection of best performing funds is a challenging task. In contrast, the loser funds exhibit persistence (see for instance Carhart, 1997 for US and Quigley and Sinquefield, 2000, for the UK).

Cuthbertson, Nitzsche and O'Sullivan (2010) provide a detailed review of literature examining mutual fund performance and performance persistence. Their review reveals the presence of data snooping bias (Sullivan et al. 1999, 2001) and estimation errors (Kothari and Warner, 2001) in the standard models for alpha estimation. In this paper, we provide a critical review of the more recent literature highlighting the shortcomings of the standard models for unbiased performance measurement. The models we review are all augmented/extended versions of the standard factor models. They are constructed with an aim of either 1) more accurate asset pricing by accounting for pricing anomalies or 2) less biased fund performance measurement, either relative to a fund's benchmark or relative to their peer group.

When it comes to anomalies, beyond those accounted for in the standard factor models (size, value and momentum), a significant strand of literature focuses on

[1] PWC 'Asset & Wealth Management Revolution: Embracing Exponential Change' report (published Oct 2017, accessed May 2018): https://www.pwc.com/gx/en/asset-management/asset-management-insights/assets/awm-revolution-full-report-final.pdf
[2] https://www.statista.com/statistics/255590/number-of-mutual-fund-companies-in-the-united-states/

adding ‘missing’ factors to the standard models aiming to explain those anomalies. The number of potential factors identified in the literature is vast (Harvey et al., 2016, Hou et al., 2015, 2017). Our review of the key recent papers shows that even though additional factors contribute to a better fit of the model, none of them can explain all anomalies in the asset-pricing context. Moreover, overall perception of the mutual fund performance and performance persistence by and large remains unchanged when the ‘missing factors’ are added. Further, the most recent academic literature raises concerns on the exploding number of the priced factors proposed and claim that some published studies are the result of statistical biases (McLean and Pontiff, 2016) and data mining (Hsu et al., 2015).

Based on the above, the main contribution of our review is centred on the most recent modifications of the standard factor models that lead to a less biased mutual fund performance evaluation by accounting for the issues associated with the presence of non-zero alphas in benchmark indices and the statistical properties of the Fama-French factor (portfolio) construction (see Cremers et. al, 2012). We continue the discussion raised in Cuthbertson et al. (2010) that both the benchmark error and evaluation of fund performance without a benchmark as a reference may cause problems. First, a fund may select a benchmark that does not match their objectives (Sensoy, 2009). Thus, if a benchmark selected for a small cap fund is a broad market index, given the evidence on outperformance of small cap stocks in the long run, it is possible that such fund will outperform its benchmark. Second, a fund may select an adequate benchmark that matches their objectives, but the benchmark itself may have a positive alpha – a phenomenon indicating a bias in construction of Fama-French risk factors (as discussed in Cremers et al., 2012)). In that case, a manager making no active bets on that benchmark will appear to have skill in the standard factor models, simply by replicating the benchmark. This is of particular interest to investors, who by and large refer to funds’ prospectus benchmarks as reference points when gauging fund performance. In addition to the evidence from the literature, the lack of clarity surrounding funds’ benchmark choices and objectives has been emphasised in the recent FCA[3] report on the UK asset management industry[4]. So not only that the

[3] Financial Conduct Authority
[4] FCA report ‘Asset Management Market Study’(published June 2017, accessed May 2018): https://www.fca.org.uk/publication/market-studies/ms15-2-3.pdf

standard models provide alphas that are not accounting for alphas embedded in the passive indices used as benchmarks; but the fund's choice of benchmark is often questionable. Hence, in this paper, we review both of those issues: the literature that highlights and/or addresses the biases in the construction of standard risk factors that lead to non-zero benchmark alphas and the literature that offers most recent suggestions for more adequate benchmarking of mutual fund performance.

Overall, the review in this paper focuses on two aspects of fund performance measurement recently challenged in literature and the evidence arising from those. First, we examine whether the models centred on mitigating the errors of the standard three- and four-factor models - in particular the 'missing factor', lead to 'winner factor/s' that account for all anomalies highlighted in the pricing literature and alter our view on mutual fund performance. Second, we evaluate the new strand of models revealing fund performance relative to some benchmark. In the latter, we first focus on models accounting for non-zero benchmark alphas and evaluate evidence on benchmark-adjusted performance of mutual funds. Then, we review the issues associated with fund's benchmark selection and evaluate the evidence from the recent addition to the factor model family – a model enabling peer-group-adjusted performance. The review is based on US and UK studies, as majority of evidence in this area stems from those markets.

The rest of the paper is organised as follows. In Section 2 we briefly revisit the standard factor models for completeness of this study. Section 3 provides summary of the main anomalies and evidence from the most recent models dealing with 'missing variables' the standard factor models. In Section 4 we review the presence of non-zero alphas in passive benchmarks and other benchmarking issues and evaluate evidence from the literature proposing some solutions to those issues. Section 5 concludes the paper.

2. Standard Factor Models

The extensive research in 1970s and 1980s[5] as well as the findings from Fama and French (1992), document superior performance of small capitalisation and high earnings-to-price (value) stocks. This prompts the extension of the original CAPM model where only the market risk is accounted for, and the development of Fama and French (1993) three-factor model, in which the size risk and the style risk associated with high book-to-market ratio firms are added. Over the years, the three-factor model became one of the most commonly referred to models in both asset pricing and performance measurement literature. In spite of its frequent use in empirical research, a number of studies point at its inability to fully explain the cross section of equity returns and highlight the need for additional risk proxies[6]. Carhart (1997) was the first to account for one persistent anomaly - the abnormal return of momentum portfolios[7] - by adding the momentum factor to the Fama-French three-factor model. Since then, these two models have become widely accepted by academics and applied in a number of studies assessing mutual fund performance and persistence in performance.[8]

The studies on mutual fund performance utilizing these standard models by and large report that active managers do not 'add value' to investors. For instance Elton and Gruber (2011) report that the variants of multi-factor models report negative alphas ranging from -0.65% (Gruber, 1996) to -1.98% (Carhart, 1997) per annum, net of fees. Further, Barras et al. (2010) for the US and Cuthbertson et al. (2012) for the UK highlight the importance of false discovery rates for funds with significantly positive alphas, i.e. funds that were lucky rather than skilful in both the short and the long run. The false discoveries account is given in Cuthbertson et al. (2010) who state that in the US and UK around 75% of active funds generate no true alphas, 20% of funds have significant negative alphas while only up to 5% of funds can be classified as true outperformers.

[5] See for instance Basu (1975),(1977) and (1983), Reinganum (1981), Banz (1981), Levis (1989), Reinganum (1992) among many others.
[6] Many of which will be summarised later in this paper.
[7] First Introduced by Jagadeesh and Titman (1993)
[8] The equations of the Fama-French and Carhart model are in the Appendix.

The most recent literature in asset pricing mainly focuses on two issues of the standard models: 1) the persistent anomalies (beyond style, value and momentum) and standard-factor model alterations to account for those anomalies, and 2) statistical properties surrounding the construction of Fama-French factors (portfolios).

3. The extensions of standard factor models: recent take on anomalies and 'missing' factors

3.1. Anomalies

Over the years a very large number of anomalies not priced by the standard (three- and) four-factor model has been identified in the literature. This section provides a summary of key anomalies, many of which motivate the extensions of the standard factor models discussed in Section 3.2.

For example, Ang et al. (2006) explain that the traditional three-factor 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. Similar is found in Jiang et al. (2009), Ang et al. (2009), Stambaugh et al. (2015) and Malagon et al. (2018). Titman et al. (2004) demonstrate a negative abnormal relation between capital investments and returns. Also, Griffin and Lemmon (2002) and Avramov et al. (2009) document a negative cross-sectional correlation between credit risk and future stock returns. Li (2011) highlights the puzzle between positive R&D-return relation and the financial constraints-return relation.

Further, Chen et al. (2011); Fama and French (1996); Fama and French (2008); Cooper et al. (2008); Daniel and Titman (2006); Campbell et al. (2008) Chen and Zhang (2010), Gray and Johnson (2011), Avramov et al. (2013) refer to anomalies such as, positive return relationship with momentum returns and earnings surprises, negative relationship with financial distress, net stock issues and asset growth and argue that the three-factor model fails to explain those. Vassalou and Xing (2004) show that the size as well as the HML factor are proxies for default effects, but only within top quintiles of firms with the highest default risk. A phenomenon where firms with high accruals underperform those with low accruals was first documented by

Sloan (1996). The accruals anomaly is found to persist even among large liquid firms (Fama and French, 2008). Mispricing due to accruals is discussed in a number of studies, see for instance Fairfield et. al (2003), Mashruwala et al. (2006), Richardson et al. (2005), Francis et al. (2005) and Hafzalla et al. (2011). However, some recent evidence documents the weakening of accruals anomaly over time (see Green et al. 2009, Hirshleifer et al., 2011) and states that expected returns to accruals based strategies are countercyclical and change over time (Wu et al. 2010).

Other anomalies from recent literature unexplained by standard pricing models include, for example, investment and investment growth (Cooper et al., 2008; Xing, 2008), return on assets (Balakrishnan et al. 2010), inventory growth (Belo and Lin, 2011) operating leverage (Novy-Marx, 2011), gross profitability (Novy-Marx, 2013), organizational capital (Eisfeldt and Papanikolaou 2013), market beta (Frazzini and Pedersen, 2014), innovations (Chen and Petkova, 2012; Cohen et al., 2013) and ex-ante skewness (Conrad et al., 2013). A comprehensive account of anomalies associated with standard factor models can be found in Hou et al. (2015). They present a list of 74 anomalies related to standard factor models, documented in academic literature, which they divide into six categories: momentum, value versus growth, investment, profitability, intangibles and trading frictions. The new study of Hou et al. (2017) extends this list even further to 447 anomalies.

3.2. Accounting for anomalies: adding the 'missing' variables

The vast number of documented anomalies and the criticism of the standard factor model's ability to explain cross-sectional returns of *all* stocks, leads us to the literature proposing augmentations to the standard factor models by adding the 'missing' variables. Harvey, Liu and Zhu (2016) review studies from the top finance, accounting and economic journals and identify 316 different factors tested in the pricing models literature. Moreover, they emphasise that the list is very likely not exhaustive. We review in this paper the evidence from the most recent studies in this area published in the top finance journals, with the main focus on the ones that not only estimate model's explanatory power, but also shed a light on portfolio/mutual fund performance.

Different definitions of liquidity have been utilized in a number of studies, to extend the standard factor models, see for instance Eckbo and Norli (2002, 2005), Pastor and Stambaugh (2003), Acharya and Pedersen (2005), and more recently, Liu (2006) and Huang et al. (2012). Those studies focus on testing ability of the proposed models with liquidity factor to explain the observed market anomalies (such as cash-flow-to price, earnings-to-price and dividend yield for instance) rather then their application in mutual fund performance measurement. For instance, Huang et al. (2012) highlight that adding liquidity to the standard Carhart model still leaves a significantly positive extreme downside risk premium unexplained. When it comes to measuring mutual fund performance, Otten and Reijnders (2012) extend the small-cap version Carhart four factor model[9] by adding the liquidity (defined as the difference in return of the low and high turnover portfolio). The model[10] is applied to the 76 UK mutual funds investing in smaller companies in the period 1992-2011. The study finds that extended model with liquidity factor yields economically and statistically significant alpha of 4.08% for small cap funds - a result which is in sharp contrast to other studies on mutual fund performance. Most other studies deal with larger or all available mutual funds in the market, but given numerous evidence on outperformance of small cap stocks[11], this result does not come as a surprise. More recently, Foran and O'Sullivan (2014) add two versions of the liquidity factor to the UK Carhart four-factor model. They use four definitions of liquidity and two liquidity factor mimicking portfolios: "illiquidity level" mimicking portfolio (obtained as returns of stocks with low minus high liquidity) and, secondly, a "systematic liquidity risk" mimicking portfolio (a measure which captures commonality in liquidity across the stocks). They examine the exposure of 1141 UK mutual funds to these liquidity risks and test their liquidity-adjusted performance. The evidence from the baseline Carhart model applied to a portfolio of funds shows that UK mutual funds are underperforming with a significant alpha of -0.14%, and when illiquidity level and liquidity risk are added, the alpha remains negative and similar in magnitude (-0.16%, significant at 1%). In the cross section, they find that liquidity level, rather than

[9] In the small-cap version of Carhart model only small-cap companies from HGSC universe[9] are used to compute the risk factors
[10] The equations for the main models discussed in this section are provided in the Appendix
[11] See Levis (1985, 1989) for the UK evidence

illiquidity, and systematic liquidity risk are both positively priced. Overall, the study finds that the model with liquidity is the one with the best fit.

Moreno and Rodriguez (2009) compare baseline unconditional Carhart model with the one augmented by the co-skewness factor and find that it increases R-squared, albeit only marginally from 84% to 85% in the sample comprised of 6819 US equity mutual funds from 1962-2006. The market, size, value and momentum risk are all positively priced and of the same magnitude and significance in both the baseline four-factor and the co-skewness model; the co-skewness factor has a small negative (-0.004) impact on mutual fund returns, significant only at 10% level (t=-1.89). The alphas remain insignificant negative values when the Carhart is augmented by co-skewness across all funds and across various sub-categories of funds; overall indicating that adding co-skewness to standard factor models does not change our view of alphas.

To account for financial distress, Ferguson and Shockley (2003) redefine the market portfolio to include debt instruments and find that adding relative leverage (debt-to-equity ratio) and relative distress (Altman's Z) to the three factor model increases the R-squared in the cross section of returns from 67% in the three-factor to 81% in the extended model. In spite of positive contribution of leverage measures in explaining the cross section of returns, they do not have notable impact on the intercept in the time series of returns of 25 Fama-French portfolios: alphas of the CAPM, the three-factor Fama-French and the extended model are of similar magnitude (mainly negative and significant); implying that leverage and distress factors are unlikely to shed a different light to a performance in a time series to that of the standard three factor model.

Jordan and Riley (2015) add a volatility factor representing the difference between the returns of low and high volatility stocks to the standard four-factor model. They report that inclusion of volatility factor reduces annual alphas for both low and high volatility mutual funds from around 5% (in a standard Carhart model) to insignificant 0.36% in the five-factor model including volatility factor, implying that extended model is more useful in capturing fund risks.

Chen, Novy-Marx and Zhang (2011) highlight the role of investment in explaining the cross-section of returns and propose a new three factor model comprised of a market factor, investment and profitability (return on assets) factor. Further, Novy-Marx (2013) argues that gross profitability is able to explain most earnings-related anomalies, as well as a large number of other anomalies. The study also suggests to alter Carhart four-factor model by employing the market and industry-adjusted value, momentum, and profitability factors. These two studies motivate one of the most notable recent contributions to the asset pricing literature – the Fama and French (2015) five-factor model, in which investment and profitability factors are added to the standard three-factor model. The notion underlying the additional factors is that firms with higher operating profitability and lower growth in total assets have above average returns. The five-factor model regressions for value weighted portfolios sorted on size and book-to-market, reveal that when compared to the standard three factor model, alphas remain significantly negative for small-growth portfolios and significantly positive for the small-value portfolios and large-growth portfolios. The largest change in the magnitude of alphas is an increase of 20 basis point per month for extreme small-cap growth portfolio (not accompanied by a change in sign or significance). Dimensional Fund Advisors research[12], which assessed the performance of actively managed mutual funds by employing Fama and French five factor model, shows that a sample of 3870 active funds in the period 1984-2015 underperforms by 0.08% per month, an amount similar to their fees. They also report that only 2.4% of funds had significantly positive alphas[13]– the results very much in line with the Fama and French (2010) finding based on the standard three-factor model.

Fama and French (2016) identify some limitations of the five-factor model when tested on the range of anomalies that remain unexplained under the standard three factor model. When testing the net share issues and volatility anomalies, the returns of portfolios in the smaller *size* quintiles (microcaps) and in the highest quintiles of share issues and volatility remain unexplained by the five-factor model. Further, accruals anomaly still persists in the five-factor model (see also

[12] http://www.fifthsetinvestment.com/using-famafrench-five-factor-model-assess-actively-managed-fund-performance/, Accessed 20th September 2018.
[13] http://novawealth.net/2018/07/18/persistent-fund-alpha-and-active-manager-skill/, Accessed 20th September 2018.

Similai, 2016 and Ball et al., 2016). The model also shows poor performance for portfolios formed on momentum, so adding momentum as the sixth factor in the model improves its explanatory power. The questionable absence of momentum in the five factor model was also highlighted in Blitz et al. (2018), who provide detailed criticism of the five-factor model and emphasize that although this new model is likely to become a new benchmark for asset pricing studies it is unlikely to bring the asset pricing debate to a consensus.

Stambaugh and Yuan (2017) attempt to tackle the issue off 'no consensus' or 'no model fits all' anomalies/stocks. They argue that anomalies reflect general mispricing and that mispricing has common components across stocks. With that in mind, they construct two 'mispricing' factors, by averaging anomaly rankings within the set of 11 anomalies examined by Stambaugh, Yu, and Yuan (2012, 2014, 2015). The set of anomalies is then divided into two clusters to form factors, with each cluster being formed based on anomalies' similarities. The results show that the four-factor model with the 'mispricing factor' explains better the set of 11 anomalies than the Fama and French (2015). Given that Hou et al. (2015) identify 74 anomalies, this still leaves 85% of these anomalies not tested here.

The recent study of Hou et al., (2017) argues that the q-factor model comprised of the market, size (market equity), investment (investment-to-assets) and profitability (return on equity) factors, proposed in Hou et al. (2015), outperforms the standard three-, four -and Fama-French five-factor models. They claim that it provides the lowest average magnitude of (and the lowest number of significant) high-minus-low alphas among all the models. Their model is tested across 161 significant anomalies, with the average magnitude of the high-minus-low decile portfolio alphas of 0.26% per month, in contrast to 0.36% in both the Carhart and the five-factor model. 46 significant high-minus-low alphas are documented for the q-factor model, versus 83 in the five-factor and 94 in the Carhart models, implying that their model is capable of pricing more anomalies than the standard models.

Other factors tested in the literature include long-run stockholder consumption risk (Malloy et al., 2009), debt capacity (Hahn and Lee, 2009), market segmentation (Menzly and Ozbas. 2010), cash holdings and risk (Palazzo, 2012), financial intermediary's wealth (Adrian et al., 2012), default and debt structure (Valta, 2016), among others. For the complete list of additional factor models proposed see Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2017).

It is clear from the review of recent literature that not any single one of the augmented factor models will be successful in explaining all anomalies, or be the best fit for all stocks. Another recent strand of new literature raises concerns over this exploding number of (priced) factors, so-called "zoo of factors" (see for instance Cochrane, 2011). Such a large set of potential factors may cause multiple comparison problems (Fama and French, 2018), lack theoretical motivation (Hou et al., 2015, Blitz et al., 2018), and represent "simply noise" stemming from data mining (Hou et al., 2015). McLean and Pontiff (2016) show that some studies suffer from statistical biases. The aforementioned criticism opens a new waive for future research aiming to propose models/solutions which identify an actual return factor in the plethora of factors. Thus, Feng et al., (2017) suggest a model-selection method aiming to bring more discipline/clarity on the set of factors recently discovered in the literature, Harvey et al., (2016) propose a framework that allows for multiple tests and derive recommended statistical significance levels for current research in asset pricing, while Fama and French (2018), propose ranking the alternative asset pricing models in terms of maximum squared Sharpe ratio for factors in a model.

The discussion in this section shows by and large that the additional factors provide a better fit (albeit sometimes only marginally), or explain some (but not all) pricing anomalies in the asset pricing context. What is more, the new models do not make a strong contribution to the literature on mutual fund performance. The evidence by and large still points towards portfolio/mutual fund underperformance even when the new models are used – as documented by the standard three-and four factor models. Therefore, even if the 'best' factor(s) for the new model is identified, its contribution is likely to be limited to asset pricing literature. Further, a strand of new literature points at the arbitrary nature and biases associated with the construction of standard factors, which leads to the presence of non-zero alphas in passive benchmark indices used in performance measurement. We therefore turn our discussion to the models dealing with the non-zero benchmark alphas.

4. Improving performance measurement: dealing with non-zero benchmark alphas

All the augmented models discussed in the Section 3 offer an extension of the standard Fama and French or Carhart models. However, Cremers et al., (2012) provide evidence suggesting that standard factor models suffer from biases and question the arbitrary nature of Fama-French factor construction method. Authors explain that size factor assigns disproportionate weight to value stocks, so this leads to a positive correlation in the SMB and HML betas of cap-weighted portfolios. Similarly, the value factor assigns disproportionate weight to small-cap stocks, therefore, it exaggerates the returns on the SMB factor. Overall, it results in an underweighting of small value stocks in the benchmark for large-cap portfolios and an overweighting of small value stocks in the benchmark for small-cap portfolios. So, this leads to a positive alpha for large stocks and a negative alpha for small stocks. The study explains that such biases in portfolio construction manifest themselves through the presence of non-zero alphas in passive benchmark indices. Hence, if a passive benchmark generates positive and significant Fama-French-Carhart alpha, a fund claiming to be an active fund can generate significant alpha by simply replicating that index. Alternatively, a fund's performance may be underestimated in the periods when its self-reported benchmark index generates negative alphas.

Several papers contribute to this discussion of non-zero benchmark alphas. Cremers et al. (2012) show that standard benchmark models produce economically and statistically significant non-zero alphas even for passive benchmark indices. It documents 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. Similarly, Matallin-Saez (2007) investigates the Russell indexes (General, Growth and Values indexes of Russell 3000, Russell 2500, Russell 2000 Russell 1000, Russell Midcap, Russell Top 200) over the period June 1995 to December 2004; reporting that all the value indexes generate positive Jensen's alpha, the highest one being that of the Russell 2500 Value Index (7.5% p.a). Chan, Dimmock and Lakonishok (2009) document a significant Fama-French alpha of -4.74% for Russell 2000 Growth index over a 13 year sample period. In other markets than US, persistently negative alphas of FTSE 100 Index was observed by Mateus, Mateus and Todorovic (2016) for the period from 1992 to 2013.

According to Chen and Knez, (1996) if the performance estimation model is correct, a passive benchmark index should not generate abnormal return. The fact that alphas are present in passive indices that serve as benchmarks for the majority of equity mutual funds, highlights the need for an unbiased pricing model that would account for the benchmark alphas. Thus, in the next section we discuss a new strand of academic literature, which proposes several augmentations of the standard factor models accounting for non-zero benchmark alpha, providing insight into fund's performance *relative* to a benchmark index.

4.1.Models accounting for non-zero benchmark alphas

To create a model that eliminates alpha in a passive index, Cremers Petajisto Zitzewitz (2012) propose the redesign of the factors used in US equity mutual fund performance evaluation in three aspects: 1) change the market portfolio to include US equities only; 2) replace equally weighted by value weighted SMB and HML factors; 3) following Moor and Sercu (2006) and similarly to Fama and French (2012), Cremers et al. (2012) decompose HML factors into value premium for big, medium and small stocks separately; further, they introduce size factors that resemble more closely size categories used in the industry: mid-cap minus large-cap and small-cap minus mid cap returns, thus creating several Carhart model augmentations. Using only returns from the US market reduces S&P500 index alpha from 0.82% to 0.52% (both significant at 1%); value weighting SMB components reduces alpha to statistically insignificant 0.33%; while introducing further augmentations reduces alphas even more. Applying modified factor models in the context of mutual fund performance does not produce alphas that substantially differ from the standard Carhart model. However, modifying the model further and replacing funds excess returns with funds' benchmark adjusted returns changes the spread between small- and large-cap fund alphas from negative (large cap outperforming small, significant at 1%) to positive (small cap outperforming large, significant at 1%) by 5.07%. This finding shows that accounting for benchmark can reverse the conclusions about fund performance established by a standard Carhart model.

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 mis-state the performance. They suggest a model that amends the left-hand side of the Carhart four-factor model to account for the benchmark-adjusted return of a fund[14]. The new alpha becomes the benchmark-adjusted alpha, a measure of fund performance relative to a benchmark index, while the factor loadings are differential factor loadings between the fund and the benchmark. Using the data for 5738 US equity mutual funds in the period September 1998 to June 2012, Angelidis et al. (2013) report average alpha for all funds from the standard Carhart model of -2.11% (t-statistics -4.43%), while their benchmark-adjusted model produces alpha of -1.25% (t-statistics -2.04%). This implies that once the benchmark alpha is taken into account, the underperformance is not as poor as the original four-factor model suggests and not as statistically significant. Nevertheless, in this paper, the broad conclusion regarding US equity mutual fund (under)performance remains robust, as per Carhart model.

In the same spirit, Mateus, Mateus and Todorovic (2016) apply Angelidis et al. (2013) alpha estimation methodology to 887 UK equity mutual funds in the period 1992-2013 and adjust fund alphas for a bias inflicted by a performance of the FTSE 100 benchmark index. Note that during the study period FTSE 100 exhibits negative alphas. Their study shows that after accounting for negative alphas of the FTSE 100, the fund performance is better than originally implied by a standard three- or four-factor model. The benchmark adjusted alpha in the full sample period improves by 127bps per year (significant at 1%). More importantly, the alpha changes from negative and significant in the Carhart model to positive and significant in the benchmark-adjusted model, a conclusion confirmed in a number of sub-periods in this study. The inverse in the results is driven by large negative benchmark's Carhart alphas documented in Mateus et al. (2016). Thus, the evidence shows that the larger the magnitude of the benchmark's Carhart alpha, the more significant the change in fund's benchmark-adjusted alpha relative to the standard Carhart alpha will be. This implies that although a fund may underperform versus the standard Carhart factors– it can still outperform *relative* to its benchmark, e.g. deliver positive benchmark-

[14] For equation, please refer to the Appendix

adjusted alpha. In contrast, in times of strong benchmark outperformance, this method helps reveal the funds that even though have generated significant positive Carhart alphas in fact fail to beat their benchmark. These findings show that the role of the benchmark in establishing the accurate, relative fund performance is important and, in some cases, can invert perceptions of investors about fund (under)performance.

Similarly, Chinthalapati, Mateus and Todorovic (2017) tackle the issue of non-zero alphas in passive benchmark indices by proposing an optimisation algorithm[15] that calculates minor fixed adjustments that should be added to the time series of the Carhart's four factors. Adjusting the factors ensures a zero alpha for any chosen self-designated benchmark index of a mutual fund, without making any other change in the model parameters. The 'adjusted Carhart factors' are then used to estimate a mutual fund's 'adjusted alpha'. The advantage of the model proposed is that it mechanically solves the issue of Fama and French factors' misspecification highlighted in the previous literature, while keeping the factor loadings and R-squared unchanged. It eliminates the issue of "the search for new factors", which may lead to unreliable statistical inferences (on this discussion see for instance Harvey et al., 2016 and Fama and French, 2018). The proposed algorithm can be applied to the three- and five-factor model as well, and to any single benchmark index. They test their novel optimisation methodology using a sample of 1383 active and tracker US equity mutual funds reporting the S&P 500 index as their prospectus benchmark, which exhibits slight outperformance over their full sample period. Subperiods in which the index has had largest non-zero alphas coincide with periods where discrepancy between Carhart and adjusted Carhart alphas is the greatest. In the full sample period, the alphas estimated with adjusted Carhart factors for mutual funds reporting S&P500 as a benchmark are lower than the standard Carhart alphas of tracker funds (by 43bp) and active funds (by 40bp) in the full sample period. The discrepancy in alphas in this study is overall small as the alpha-adjustment is based on small positive S&P500 Carhart alphas, documented for the US. Thus, the results do not change overall inferences on the US funds' underperformance; however, highlight, that the benchmark adjusted fund's performance is in fact worse that initially documented by the standard Carhart model.

[15] Available from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2581737

Overall, the literature in this section shows that accounting for benchmarks in mutual fund performance measurement matters. Positive and significant benchmark alphas imply lower benchmark-adjusted alphas compared to Carhart alphas of a fund. So a fund with significantly positive alpha in the Carhart model, can exhibit no outperformance or even negative alphas when adjusted for benchmark. The more negative alpha of the fund's benchmark index is, the more likely it is that the fund's alphas will revert from significantly negative in the Carhart model to positive in the benchmark-adjusted models; implying that even though the fund does not perform well given the Carhart risk parameters, it still performs better than its benchmark in times of distress.

4.2. Benchmarks not aligned with fund objectives

The augmented models discussed in the previous section require the use of performance benchmark, typically a fund's self-reported (self-designated) prospectus benchmark. Hence, a fund's choice of benchmark is particularly relevant to investors interested in fund's performance relative to that benchmark. In that case, the most accurate performance can be estimated when the reference benchmark selected is based on fund's holdings and objectives. It is, however, not uncommon in practice that mutual funds in the same peer-group[16] benchmark against a number of different passive indices. A new strand of literature reviewed in this section raises concerns on the content of information disclosed by the funds, and in particular, on adequacy of their prospectus benchmarks given the fund's investment style and objectives.

Sensoy (2009) finds strong evidence that self-designated benchmarks are consistently mismatched by mutual funds. Thus, results show that 31.2% of funds analysed specify a benchmark index that is less adequate compared to some other S&P or Russell size and value/growth-based benchmark ("corrected" benchmarks, which better match funds' size or value/growth characteristics, and are more correlated with funds' returns). The author highlights that funds tend to select benchmarks that are easier to 'beat', so that their outperformance generates greater fund inflows. Such mis-specified benchmarks are more common among large and high-fee funds. More

[16] By peer-group here, we refer to the funds following the same investment style as per Morningstar classification.

recently, Mateus, Mateus and Todorovic (2018a) using Morningstar style categories show that 2/3 of the funds that report S&P500 as their benchmark have risks and objectives different to those of the S&P500 index. They deploy Angleidis et al. (2013) approach and compare benchmark-adjusted alphas when that benchmark is 1) S&P500 and 2) an index having a better fit to the fund's Morningstar style category. They find that S&P500-adjusted alphas are higher than style category benchmark-adjusted alphas in 61.2% of the cases; adding to the evidence that some funds may strategically select a benchmark that is easier to 'beat'.

Similar is found by Cremers and Petajisto (2009), who provide evidence that mutual funds typically have a high proportion of holdings that differ from those in 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. More recently, 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. They also reveal that misclassified funds tend to be younger, smaller in size and charge higher expense ratios. The issue of mismatched benchmarks is present in earlier studies too. 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).

These findings send a message to investors interested in fund performance relative to the benchmark: the adequacy of the selected benchmark for the fund and its match to fund objectives needs to be assessed before engaging in any performance evaluation relative to that benchmark. For regulators, the findings give a message that tighter regulation on benchmark selection is needed in the mutual fund industry, such as compulsory benchmark requirement for funds that do not report any benchmark and disclosure of benchmark selection methodology for funds that do report one.

4.3.Potential solutions for the mis-matched benchmarks problem

Several studies propose some solutions to the problem of mis-matched benchmarks. One solution can be found in the Sharpe (1992) style analysis study. Although the original purpose of this model was to decompose portfolio returns into asset classes a portfolio invests in, and separate the returns resulting from asset allocation and security selection; the method can give useful indication of what would be an adequate mutual fund benchmark. For instance, if the style analysis reveals that the 65% of mutual fund's return stems from allocation to a value style index and 35% from allocation to a small-cap index, then a suitable benchmark for that fund is value-small cap index. Sharpe style analysis is based on historical returns (rather than holdings) data and it is not very successful in capturing sudden style drifts in a fund[17]. Rather than looking at returns of funds, Daniel, Grinblatt, Titman and Wermers (1997) use characteristics-based benchmarks, i.e. benchmarks matching characteristics of stocks that mutual funds hold in their portfolios. Conceptually, this study corresponds to Brinson, Hood and Beebower (1986) performance attribution analysis where portfolio excess return is separated into asset (factor) allocation and security selection components. Daniel et al. (1997) method cannot be applied unless the complete information on fund holdings is available, which is not the case for many markets. A more recent study by Chan et al. (2009), proposes to estimate performance relative to 1) characteristic-matched benchmarks (constructed based on

[17] https://web.stanford.edu/~wfsharpe/art/fa/fa.htm Accessed 28th September 2018.

size-conditional book-to-market sorts, quarterly size-conditional book-to-market sorts, size-conditional composite value/growth indicator approach) and 2) 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.

These studies show that the choice of performance evaluation methodology and the choice of a referred benchmark impact the performance, as changing the benchmark or the evaluation method incurs a change in fund alphas. It is important to note here that a change in the value of alpha does not necessarily change the ranking of the fund within a group - an issue not discussed in the aforementioned studies - but of essence to investors trying to select the best fund within a group (say, the best small-cap fund). If the entire peer-group of funds benchmarks against the *same* index, then, adjusting fund alphas for benchmark alphas as in Angelidis et al. (2013) or Chinthalapati et al. (2017) will change the value of fund alpha (alphas for all funds decrease if a benchmark is outperforming and vice versa); but *not* the fund's ranking position in the peer-group. However, if such a benchmark does not match objectives of a fund, its peer group ranking is biased, simply because a fund may strategically choose a benchmark that is easy to beat. Further, if funds within the same peer group benchmark against different passive indices, then their benchmark-adjusted alphas are not directly comparable.

A recent study of Hunter at al. (2014) proposes a novel methodology designed for peer-group setting. The method provides solution to the issue of non-zero benchmark alphas generated by standard factor models and offers an approach to eliminate biases inflicted in inaccurate fund self-reported benchmarks. It is based on the idea that prior to making a choice on which fund to invest in, investors identify the group of funds, which according to the investment objectives suits them best. Then, they select the fund within that peer-group with the best past performance. Hunter et al. (2014) emphasise that the manager's true skill/performance should exclude the commonalities in fund strategies within a peer group. Hence, the authors argue that instead of adding a number factors to the standard four-factor model to account for complex strategies within a peer-group, there should be only one factor added – a factor that accounts for peer group commonalties in idiosyncratic risk-taking and allows estimation of unique manager skills, uncorrelated to the average of the peer group. Thus, the study proposes construction of the Active Peer Benchmark (APB hereafter, viewed as a passive benchmark) as an equally-weighted portfolio of all the funds in the peer group. The sum of the estimated four-factor alpha of the APB benchmark and the error term from that four-factor model[18], represents the additional APB factor (the fifth-factor) in the proposed augmented Carhart model. The alpha in that augmented model, the APB model, is the peer-group adjusted alpha. Thereby, if a fund manager has skills that are above common strategies used within the reference group, the APB adjusted alpha in the new APB adjusted model will be positive and significant. The APB benchmark eliminates the need to use funds (potentially biased) self-reported benchmark.

When it comes to contribution of their model to asset pricing, Hunter et al. (2014) show that adding commonalities in fund strategies (APB related factor) to the Carhart model results in a greater explanatory power of their model. While one may argue that the APB method is another example of adding yet another factor to the Carhart four-factor model, we want to highlight that the purpose of the inclusion of the active peer benchmark is different from other studies searching for a 'missing' priced factor while targeting to solve the anomaly puzzle discussed in the Section 3. The methodology discussed in Hunter et al. (2014) aims to facilitate/improve investor selection of top performing funds within a comparable (peer) group of funds and identifies managerial skills that exceed the average skill of the group. Most importantly, in contrast to other methodologies suggesting the additional priced factors, it is only the model that tackles the issue of non-zero benchmark alphas and offers a solution to the benchmark selection problem by using a relevant peer-group as a benchmark. This means that the choice of the comparable (peer) group is of importance here and should be based on some broadly accepted standard. For instance, choosing a peer-group by Carhart factor loadings would not be an adequate solution as it leads to misclassification of some, particularly large cap funds due to arbitrary factor construction (Cremers et al, 2012, and Chen and Basset, 2014). Better alternative for investors is to use standardised peer-group classification of funds such as CRSP classification codes or a widely accepted industry performance monitoring groupings, such as Morningstar, where equity funds are assigned to a style category

[18] Appendix shows equation of the Hunter et al. (2014) APB model.

based on their historical holdings. Investors should, of course, monitor for any changes in fund peer-group allocation over time, whichever standardised peer-group data source they use.

Empirical evidence shows that APB adjusted model is found to be better in identifying the top and bottom performing funds/managers within the peer-group than the standard Carhart model. Hunter et al. (2014) findings hold for both US equity and bond funds in the period 1980-2010. This methodology was also tested in Mateus, Mateus and Todorovic (2018b) on a sample of 817 UK long-only active equity mutual funds allocated to nine Morningstar based peer-groups for the period 1992-2016. The study documents that APB adjusted model is able to identify the top performing funds versus the peer group, which continue to perform well one-year-ahead.

The studies in this section differ from the ones in Section 3 as their purpose is to contribute to performance measurement, not to asset pricing. All the benchmark-adjusted and peer-group adjusted models assume that the most unbiased fund performance *can* be estimated based on the widely accepted standard Fama-French-Carhart risk factors if fund performance is accounted for the funds' self-declared benchmark (Angelidis et al., 2013, Chinthalapati et al., 2017) or peer-group benchmark (Hunter et al, 2014) as it eliminates the issues of arbitrary factor construction and a potentially mismatched reference benchmark. These new models are of use to investors aiming to identify funds *relative* performance, i.e. benchmark-adjusted or peer-group adjusted performance. Accounting for positive benchmark alphas can help disclose closet index funds that claim to be active but in fact perform in line with index; while accounting for negative index performance may reveal funds that do better than the benchmark at the time of distress. It is also important for investors to assess performance relative to the peer –group, as a fund may have a positive Carhart alpha but actually be worse than the average in the peer-group. Hence, the literature discussed in this section provides important evidence on the role of the benchmark in establishing the relative fund performance and is important for future research. The discussed above new models bring modifications to the standard alphas that shed new light on the *relative* performance of funds. Nevertheless, investors should bear in mind that these models will only be useful if the benchmark for relative performance is defined well, as discussed previously in this section.

5. Conclusions

Standard factor models such as Fama and French (1993) and Carhart (1997) have long been accepted as a norm in academic studies assessing mutual fund performance measurement. Cuthbertson's et al. (2010) review of literature emphasises on the application of these models on performance, persistence in performance of mutual funds and false discoveries (fund manager's luck vs. skill). More recently, Mason et al. (2016) focus on identifying new trends in investment management and performance following the discussion on skill versus luck, managerial characteristics and incentives. The main contribution of this paper is that it reviews the evidence from the new, augmented asset pricing models that enable 1) more accurate asset pricing by accounting for pricing anomalies or 2) less biased relative performance measurement, either relative to the fund's benchmark or relative to their peer group.

A large number of anomalies, not priced by the standard factor models, have been identified in the literature; Hou et al. (2017) lists as many as 447. With this in mind, numerous studies extend the standard factor models by adding 'missing variables' in an attempt to improve the model's fit and price some of the anomalies. In this review we focus on whether these extensions of the standard models shed a new light to mutual fund performance. We find that these extended models by and large do provide a better fit to a cross section of equity returns but they are still far from determining which *one* of the plethora of additional factors suggested in the literature is the 'winner'. In the studies that provide some evidence on mutual fund performance, adding the new factors only confirms underperformance of funds, as determined by the standard three- and four-factor models. Moreover, recent academic studies claim that the majority of the extended models proposed aiming to explain anomalies are results of data mining (Hou et al., 2015; 2017) and lack theoretical motivation (Blitz et al., 2018). Hence, the newest trend in literature on asset pricing is most recently evolving in the direction of Fama and French (2018), where metrics for ranking alternative models are discussed.

Our main focus in this review is evidence from the new strand of asset pricing models that amend mutual fund performance relative to the benchmark index and relative to their peer-group. The literature provides strong evidence of presence of non-zero alphas in benchmark indices commonly used by funds as their prospectus benchmarks. Augmenting standard Carhart model to accounting for alpha of the benchmark index can change the traditional view of mutual fund performance. Thus, in the periods of index outperformance, it is possible to identify an active fund that generates positive alpha in the standard Carhart model but actually only replicates benchmark returns. Vice versa, at times of significant index underperformance, a fund generating negative Carhart alpha can still have a positive benchmark-adjusted alpha or have better skill than the peer-group even if the entire peer-group underperforms. By and large the evidence shows that the relationship between the benchmark's Carhart alpha and a fund's benchmark-adjusted alpha is inverse. If the benchmark's Carhart alpha is of the same sign but of greater magnitude than the fund's Carhart alpha, the fund's benchmark-adjusted alpha will have an inverse sign relative to its Carhart alpha, thus changing investor's perception of fund performance.

While the academic studies in relative performance measurement are easily applicable and give opportunity to investors to directly adjust a fund's performance for that of their benchmark, they indirectly raise a doubt about which benchmark to utilise in the model. Typically, investor will be assessing performance against a fund's self-declared prospectus benchmark or against their peer-group, whereby examining how a fund fares against other funds with similar objectives. We recognise that recent literature raises the issue of prospectus benchmarks not being aligned with fund objectives, hence inflicting a bias in benchmark-adjusted performance measurement. We are of the view that, in practice, funds need to state in their prospectus how and why a particular passive index is selected as a benchmark, as well as provide detailed fund objectives. In the UK, the FCA is making a step in the right direction in this area, where more disclosure from funds regarding benchmarks and objectives will be required. That will help re-assure investors that the prospectus benchmark is, indeed, the most appropriate one, prompting a more widespread use of benchmark-adjusted models such as Angelidis et al. (2013) and Chinthalapati et al. (2017) in performance evaluation. Alternatively, with the data sources getting more comprehensive and advances in computer technology, building characteristics-based benchmarks that change over time to capture style drifts in a fund is another way forward for more accurate benchmark-adjusted performance.

For funds declaring the same passive index as a benchmark, the inclusion of that benchmark in performance measurement changes the value of alpha but not the ranking of the fund. Therefore, a fund may outperform their benchmark but still be at the bottom of their peer-group (if all other funds do better than that benchmark). Investors interested in selecting the best funds within a peer-group should construct Active Peer Group benchmark as in Hunter et al. (2014). The model is shown to have better ability to select winner funds than the standard Carhart model. The composition of the peer-groups may have been debatable in the past, but using standardised peer-group classifications, such as those provided by Morningstar, are nowadays widely accepted peer-group classifications in practice.

Given that standard factor models, Fama-French three- and Carhart four-factors have not been widely used in mutual fund performance measurement in practice, the new models in relative fund performance that better reflect what is of importance to investors, i.e. performance against the benchmark and the peer-group, are well suited for application in the industry.

At present, these new developments in fund performance evaluation are centred on equity mutual funds and models that are appropriate for pricing the cross-section of equity returns. With a growing demand for multi-asset funds[19] and the inclusion of alternative asset classes in more traditional portfolios, the need for a more comprehensive pricing models better suited to those assets arises. Further, the models augmented with additional factors, the benchmark-adjusted and peer-group adjusted models discussed in this review are all unconditional models. Future work in this area may revert attention to conditional benchmark-adjusted and peer-group adjusted performance measurement.

[19] https://www.investorschronicle.co.uk/funds-news/2018/01/25/multi-asset-fund-launches-capitalise-on-strong-demand/ (Accessed 22nd May 2018)

References

Adrian, T., Etula, E. and Muir, T., 2014. Financial intermediaries and the cross-section of asset returns. The Journal of Finance, 69(6), pp.2557-2596.

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

Ang, A., Hodrick, R.J., Xing, Y. and Zhang, X., 2009. High idiosyncratic volatility and low returns: International and further US evidence. Journal of Financial Economics, 91(1), pp.1-23.

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

Acharya, V.V., and L.H. Pedersen, 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375–410.

Avramov, D., Chordia, T., Jostova, G. and Philipov, A., 2009. Credit rating and the cross-section of stock returns, Journal of Financial Markets, 12:3, pp.469-499.

Avramov, D., Chordia, T., Jostova, G. and Philipov, A., 2013. Anomalies and financial distress. Journal of Financial Economics, 108:1, pp.139-159.

Balakrishnan, K., E. Bartov, and L. Faurel. 2010. Post loss/profit announcement drift. Journal of Accounting and Economics 50:20–41.

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.

Banz, R. W., 1981. “The relationship between return and market value of common stocks”, Journal of Financial Economics, 9:1, 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, 65(1), pp.179-216.

Basu, S., 1975, The information content of price-earnings ratios, Financial Management, 4, pp.53-64.

Basu, S., 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of efficient market hypothesis, Journal of Finance, 32, 663-81.

Basu, S., 1983, The relationship between earnings' yield, market value and return for NYSE common stocks, Journal of Financial Economics, 12:1, pp. 129-156

Belo, F., and X. Lin. 2011. The inventory growth spread. Review of Financial Studies 25, pp. 278–313

Berk, J.B. and Van Binsbergen, J.H., 2012. Measuring managerial skill in the mutual fund industry (No. w18184). National Bureau of Economic Research.

Blitz, D., Hanauer, M.X., Vidojevic, M. and van Vliet, P., 2018. Five Concerns with the Five-Factor Model The Journal of Portfolio Management Quantitative Special Issue 2018, 44 (4) pp. 71-78

Brinson, G.P., Hood, L.R. and Beebower, G.L., 1986. Determinants of portfolio performance. Financial Analysts Journal, pp.39-44.

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

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

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, pp. 4553-4599.

Chen, L., Novy-Marx, R. and Zhang L., 2011, An Alternative Three-Factor Model. Available at SSRN: https://ssrn.com/abstract=1418117

Chen, L. and Zhang L., 2010, A Better Three-Factor Model That Explains More Anomalies. The Journal of Finance, 65:2, pp.563-594

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

Chen, Z. and Petkova, R., 2012. Does idiosyncratic volatility proxy for risk exposure?. The Review of Financial Studies, 25(9), pp.2745-2787.

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.

Cohen, L., K. Diether, and C. Malloy. 2013. Misvaluing innovation. Review of Financial Studies 26: pp. 635–66.

Conrad, J., R. F. Dittmar, and E. Ghysels. 2013. Ex ante skewness and expected stock returns. Journal of Finance 68: pp. 85–124.

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. and O'Sullivan, N., 2012. False Discoveries in UK Mutual Fund Performance. European Financial Management, 18(3), pp. 444–463

Cochrane, J.H., 2011. Presidential address: Discount rates. The Journal of finance, 66(4), pp.1047-1108.

Cooper, M., H. Gulen, and M. Schill. 2008. Asset growth and the cross-section of stock returns. Journal of Finance 63:1609–52.

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

Daniel, K. and Titman, S., 2006, Market Reactions to Tangible and Intangible Information, Journal of Finance, 61(4), pp. 1605–1643

diBartolomeo D. and Witkowski, E., 1997. Mutual fund misclassification: Evidence based on style analysis. Financial Analysts Journal, 53:5, pp.32-43.

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

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

Eisfeldt, A., and D. Papanikolaou. 2013. Organizational capital and the cross-section of expected returns. Journal of Finance 68:1365–1406.

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

Elton, E., Gruber, M. 2011. Mutual funds. Working paper. https://ssrn.com/abstract=2088418

Evgeniou, T., de Fortuny, E.J., Nassuphis, N. and Vermaelen, T., 2018. Volatility and the buyback anomaly. Journal of Corporate Finance, 49:C, pp. 32-53.

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

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 French, K.R., 2010. "Luck Versus Skill in the Cross-Section of Mutual Fund Returns." Journal of Finance, 65(5): pp.1915-47.

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

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

Fama, E., French, K., 2018, Choosing factors, Journal of Financial Economics, 128:2, pp. 234-252

Feng, G., Giglio, S. and Xiu, D., 2017. Taming the factor zoo. Working paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2934020

Ferguson, M.F., and R.L. Shockley, 2003. Equilibrium anomalies, Journal of Finance 58:6, pp. 2549–2580.

Francis, J., R. La Fond, P. Olsson, and K. Schipper. 2005. The market price of accruals quality. Journal of Accounting and Economics 39:295–327.

Frazzini, A., and L. Pedersen. 2014. Betting against beta. Journal of Financial Economics 111, pp. 1–25

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

Gray, P. and Johnson, J., 2011. The relationship between asset growth and the cross-section of stock returns. Journal of Banking & Finance, 35(3), pp.670-680.

Green, J., Hand, J.R. and Soliman, M.T., 2011. Going, going, gone? The apparent demise of the accruals anomaly. Management Science, 57(5), pp.797-816.

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

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

Harvey, C. R., Liu, Y. and H. Zhu, 2016. Editor's Choice … and the Cross-Section of Expected Returns, Review of Financial Studies, 29: 1, pp. 5-68

Hafzalla, N., R. Lundholm, and E. Van Winkle. 2011. Percent accruals. The Accounting Review 86, pp. 209–36

Hahn, J., and H. Lee. 2009. Financial constraints, debt capacity, and the cross-section of stock returns. Journal of Finance 64:891–921.

Hirshleifer, D., Teoh, S., Yu, J., 2011. Short arbitrage, return asymmetry, and the accrual anomaly. Review of Financial Studies 24 (7), 2429–2461.

Hou, K., Xue, C. and Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies, 28(3), pp.650-705.

Hou, K., Xue, C. and Zhang, L., 2017. Replicating anomalies (No. w23394). National Bureau of Economic Research. Working paper https://www.nber.org/papers/w23394

Hsu, J., Kalesnik, V. and Viswanathan, V., 2015. A framework for assessing factors and implementing smart beta strategies. The Journal of Index Investing, 6(1), p.89.

Huang, W., Q. Liu, S. G. Rhee, and F. Wu. 2012. Extreme downside risk and expected stock returns. Journal of Banking & Finance 36:pp. 1492–502.

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.

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

Jiang, G.J., Xu, D. and Yao, T., 2009. The information content of idiosyncratic volatility. Journal of Financial and Quantitative Analysis, 44(1), pp.1-28.

Jordan, Bradford and Riley, Timothy B., (2015), Volatility and mutual fund manager skill, Journal of Financial Economics, 118: 2, pp. 289-298.

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 61:6, pp. 2551-2595.

Kothari, S.P. and J.B. Warner, 2001, Evaluating Mutual Fund Performance, Journal of Finance 56:5: pp.1985-2010

Levis, Mario, 1985, Are small firms big performers, The Investment Analyst 76, 21-27.

Levis M., 1989, Stock market anomalies: a re-assessment based on UK evidence. Journal of Banking and Finance, 13, pp.675-696.

Li, D., 2011. Financial constraints, R&D investment, and stock returns. The Review of Financial Studies, 24(9), pp.2974-3007.

Liu, W., 2006. A Liquidity-Augmented Capital Asset Pricing Model, Journal of Financial Economics, 82 (3), pp. 631-671.

Malagon, J., Moreno, D. and Rodríguez, R., 2018, Idiosyncratic volatility, conditional liquidity and stock returns. International Review of Economics & Finance, 53, pp.118-132.

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jorgensen. 2009. Long-run stockholder consumption risk and asset returns. *Journal of Finance* 64:2427–79.

Mashruwala, C., Rajgopal, S. and Shevlin, T., 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. Journal of Accounting and Economics, 42(1-2), pp.3-33.

Mason, A., Agyei-Ampomah, S. and Skinner, F., 2016. Realism, skill, and incentives: Current and future trends in investment management and investment performance. International Review of Financial Analysis, 43, pp.31-40.

Mateus, I. B., Mateus, C. and Todorovic, N., 2016, UK equity mutual fund alphas make a comeback, International Review of Financial Analysis, 44:C, pp. 98-110.

Mateus, I. B., Mateus, C. and Todorovic, N., 2018a, The impact of benchmark choice on US mutual fund benchmark-adjusted performance and ranking, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3014010

Mateus, I. B., Mateus, C. and Todorovic, N., 2018b, UK Mutual Fund Performance Persistence with Active Peer Benchmarks, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3011829

Matallin-Saez, J.C. (2007), Portfolio Performance: Factors or Benchmarks? Applied Financial Economics, 17:14, pp.1167–1178.

Menzly, L., and O. Ozbas. 2010. Market segmentation and cross-predictability of returns. Journal of Finance 65:1555–80.

McLean, R.D. and Pontiff, J., 2016. Does academic research destroy stock return predictability?. The Journal of Finance, 71(1), pp.5-32.

Moor, L.De and Sercu, P. (2006), The small firm anomaly: US and international evidence, working paper, Katholieke Universiteit Leuven

Moreno D., Rodríguez R., 2009. The value of coskewness in mutual fund performance evaluation, Journal of Banking & Finance, 33:9, pp. 1664–1676

Novy-Marx, R., 2011. Operating leverage. Review of Finance, 15(1), pp.103-134.

Novy-Marx, R. 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, pp. 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 Stambaugh, R.F., 2003. Liquidity risk and expected stock returns, Journal of Political Economy, 111:3, pp. 642–685.

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

Palazzo, B. 2012. Cash holdings, risk, and expected returns. Journal of Financial Economics 104: pp. 162–85.

Reinganum, M. R., 1981, Misspecification of capital asset pricing: empirical anomalies based on earnings yields and market values, Journal of Financial Economics 9, pp.19-46.

Reinganum, M. R., 1992, A revival of the small firm effect, Journal of Portfolio Management, 18, Spring 1992, pp.55-62

Richardson, S., R. Sloan, M. Soliman, and I. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. Journal of Accounting and Economics 39:437–85.

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., 1992. Asset allocation: Management style and performance measurement. Journal of portfolio Management, 18(2), pp.7-19.

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.

Stambaugh, R.F., and Yuan, Y, 2017, Mispricing factors, The Review of Financial Studies, 30:4, pp. 1270–1315.

Stambaugh, R.F., Yu, J. and Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics, 104(2), pp.288-302.

Stambaugh, R.F., Yu, J. and Yuan, Y., 2014, The long of it: Odds that investor sentiment seriously predicts anomaly returns, Journal of Financial Economics, 114, pp. 613-619.

Stambaugh, R.F., Yu, J. and Yuan, Y., 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle. The Journal of Finance, 70(5), pp.1903-1948.

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

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

Titman, S., Wei, K.J. and Xie, F., 2004. Capital investments and stock returns. Journal of financial and Quantitative Analysis, 39(4), pp.677-700.

Valta, P., 2016. Strategic default, debt structure, and stock returns. Journal of Financial and Quantitative Analysis, 51(1), pp.197-229.

Vassalou, M., and Xing Y, 2004, Default risk in equity returns, Journal of Finance, 59:2, pp. 831–868.

Wermers, R. 2000. Mutual Fund Performance: An Empirical Decomposition into Stock Picking Talent, Style, Transactions Costs, and Expenses. Journal of Finance 55:4, 1655-1695.

Wu, J.G., Zhang, L. and Zhang, X., 2010. The q-theory approach to understanding the accrual anomaly. Journal of Accounting Research, 48:1, pp.177-223.

Xing, Y. 2008. Interpreting the value effect through the Q-theory: An empirical investigation. Review of Financial Studies 21:1767–95.

Appendix

Standard benchmark models	
Jensen (1968)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \varepsilon_{i,t}$ Were $(R_{it} - r_{ft})$ is excess return of mutual fund *i* over the risk free rate and $(R_{mt} - r_{ft})$ is excess return of the market, typically proxied by a benchmark index. β_{im} is the indicator of the systematic (market) risk and it shows the sensitivity of fund returns to the market returns. $\propto_i$ represents manager skill, i.e. excess return of the fund once the Market risk is accounted for. $\varepsilon_{i,t}$ is the standard error term.
Fama and French (1993)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \varepsilon_{it}$ Where $(R_{mt} - r_{ft})$ is excess return of the market, SMB represents a size factor obtained as a difference in returns between small cap and large (big) cap firms and HML is the style factor obtained as the difference in returns between the firms with the high book-to-market (value firms) and low book-to-market ratio (growth firms). $\propto_i$ represents fund's excess return after the market risk, size and style risk are taken into account.
Carhart (1997)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iWML}WML_t + \varepsilon_{it}$ Where WML is momentum factor obtained as the difference in winner (top 30% of firms with the highest 11-month returns) and loser (bottom 30% of firms with the lowest 11-month returns) returns. In this model, $\propto_i$ is the excess return of mutual fund obtained after adjusting for the market, size, style and momentum risk. The rest is as per Fama and French (1993) model.
Augmented standard models with added factors	
Novy-Marx and Zhang (2011)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{INVi}INV_t + \beta_{ROAi}ROA_t + \varepsilon_{it}$ Where Investment (INV) factor represents the difference in the returns between low and high investment firms and profitability, measured by return on assets (ROA) represents the difference in the returns of firms with high ROA and low ROA.

Novy-Marx (2013)	$R_{it} - r_{ft} = \propto_i + \beta_{iMKT} MKT_t + \beta_{iHML} HML^*{}_t + \beta_{iUMD} UMD^*{}_t + \beta_{iPMU} PMU^*{}_t + \varepsilon_{it}$ Where HML* is industry-adjusted high-minus-low factor, UMD* -industry-adjusted up-minus-down factor, PMU*- industry-adjusted profitable-minus-unprofitable factor. The findings show that the proposed model with industry-adjusted factors is able to price 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.
Hou, Xue and Zhang (2015)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{SMBi} SMB_t + \beta_{I/Ai} I/A_t + \beta_{ROEi} ROE_t + \varepsilon_{it}$ q-factor model, where size is the market equity, estimated as stock price per share times shares outstanding from CRSP; investment factor (I/A, investment-to-assets) represents the difference between the return on a portfolio of low investment stocks and the return on a portfolio of high investment stocks (I/A is the annual change in total assets divided by one-year-lagged total assets), ROE factor is the difference between the return on a portfolio of high profitability (return on equity, ROE) stocks and the return on a portfolio of low profitability stocks (RoE is income before extraordinary items divided by one-quarter-lagged book equity), SMB is from Fama and French (1993)
Otten and Reijnders (2012)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{SMBi} SMB_t + \beta_{HMLi} HML_t + \beta_{WMLi} WML_t + + \beta_{LMHi} LMH_t + \beta_{D,i} D_t + \varepsilon_{it}$ (6) where LMH represents the difference in return of the low and high turnover portfolio, while D is a dummy variable taking value of one in January and zero in all other months. The rest is as per standard Carhart (1997) model.
Foran and O'Sullivan (2014)	$R_{it} - r_{ft} = \alpha_i + \beta_M R_{mt} + \beta_S SMB_t + \beta_V HML_t + \beta_{MOM} MOM_t + + \beta_L LIQ_t + e_{t,}$ where R_{mt} is the FTSE All Share return in month t in excess of 3-month sterling denominated gilts; SMB, HML, MOM are UK size, value and momentum factors; LIQ is either the illiquidity characteristic risk or systematic liquidity risk mimicking portfolio (or both may be specified in some model estimations).
Moreno and Rodriguez (2009)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{SMBi} SMB_t + \beta_{HMLi} HML_t + \beta_{WMLi} WML_t + \beta_{CSKi} CSK_t + \varepsilon_{it}$ Where CSK is defined as the return on the assets with the most negative co-skewness minus return of the assets with highest

	positive co-skewness; all else as in the standard Fama-French three-factor model.
Hirshleifer and Jiang (2010)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{SMBi}SMB_t + \beta_{HMLi}HML_t + \beta_{UMOi}UMO_t + \varepsilon_{i,t}$ Where UMO is the difference in returns of undervalued and overvalued stocks and everything else is as per standard Fama and French (1993) model.
Jordan and Riley (2015)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iWML}WML_t + \beta_{iLVH}LVH_t + \varepsilon_{it}$ Where LVH a low vs. high volatility factor representing a difference between the returns of low and high volatility stocks; all else is the same as in Fama and French (1993).
Fama and French (2015)	$R_{it} - r_{ft} = \propto_i + \beta_{im}(R_{mt} - r_{ft}) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iWML}WML_t + \beta_{iRMW}RMW_t + \beta_{iCMA}\beta CMA_t + \varepsilon_{it}$ Where RMW is the profitability factor obtained as the return spread of the firms with robust profitability (30% of firms with highest operating profitability) and week profitability (30% least profitable firms based on operating profitability). CMA denotes the investment obtained as the return spread of firms that invest conservatively (reflected through low total asset growth) and those that invest aggressively (i.e. have a high asset growth).The rest is as in Fama and French (1993).
Stambaugh and Yuan (2017)	$R_{it} - r_{ft} = \propto_i + \beta_{iMKT}MKT_t + \beta_{iSMB}SMB_t + \beta_{iMGMT}MGMT_t + \beta_{iPERF}PERF_t + \varepsilon_{it}$ Where MKTt is the excess market return, SMBt is the size factor, and MGMTt and PERFt are the mispricing factors. MGMT is based on anomalies from the management cluster: net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment to assets; while PERF is based on anomalies from the performance cluster: distress, O-score, momentum, gross profitability, and return on assets.
Models adjusted for non-zero benchmark alphas	
Angelidis, Giamouridis and Tessaromatis (2013)	$R_{i,t} - R_{Benchmark,t} = \alpha_i^* + \beta_{i1}^*(R_{M,t} - R_{f,t}) + \beta_{i2}^*SMB_t + \beta_{i3}^*HML_t + \beta_{i4}^*WML_t + \varepsilon_{it}^*$ Where $R_{i,t} - R_{Benchmark,t}$ is the benchmark adjusted return of a

	mutual fund i in period t., α_i^* is the fund's benchmark-adjusted alpha representing the difference between fund's standard four-factor alpha and benchmark index's standard four-factor alpha. $(R_{M,t} - R_{f,t})$, SMB, HML and WML are defined in equation (3). $\beta_{i1}^*, \beta_{i2,}^* \beta_{i3,}^* \beta_{i4}^*$ represent the difference between the fund's and benchmark's Carhart betas, i.e. benchmark-adjusted betas, hence showing how much a fund portfolio over/under weights small/large or value/growth stocks relative to the self-reported benchmark index.
Chinthalapati, Mateus and Todorovic (2017)	Authors use an optimisation algorithm[20] to derive fixed (time invariant) factor adjustments f_i (i =1,2,3,4), and adjusted factors $AF_{i,t} \equiv F_{i,t} - f_i$, so that: $R_{b,t} = \alpha_b^{adj} + \sum_{i=1}^{4} \beta_{b,i}^{adj} AF_{i,t} + \varepsilon_{b,t}^{adj}$ The algorithm calculates minimal fixed adjustments to the standard Carhart's four factors that 1) ensure the estimated benchmark alpha is (close to) zero (with lowest possible t-statistic), 2) maintain the same R2 and the same factor beta coefficients (and their statistical significance) as when using the standard Carhart factors. $\alpha_k^{adj}$is the benchmark-adjusted Carhart four-factor alpha
Hunter at al. (2014)	$R_{i,t} - R_{f,t} = \alpha_{i,ADJ} + \beta_{i,M}(R_{M,t} - R_{f,t}) + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,WML}WML_t + \beta_{i,ADJ}(\widehat{\alpha_{APB,\iota}} + \widehat{e_{APB\iota,t}}) + \varepsilon_{i,ADJ,t}$ Two-stage APB model, where $\widehat{\alpha_{APB,\iota}}$, is estimated alpha of the equal-weighted active peer group that fund i belongs to and $\widehat{e_{APB\iota,t}}$ is four- factor regression residuals for the APB to the four- factor model for fund i; . $\widehat{\alpha_{APB,\iota}} + \widehat{e_{APB\iota,t}}$. is the APB adjustment factor, and the new $\alpha_{i,ADJ}$ is the fund's alpha adjusted for the peer-group benchmark, the APB.

[20]The optimisation code written in Matlab is available from the working paper version of this article, see https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2581737 .