The European Journal of Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rejf20

Mutual fund performance: the model for selecting persistent winners

Cesario Mateus, Irina B. Mateus & Natasa Todorovic

To cite this article: Cesario Mateus, Irina B. Mateus & Natasa Todorovic (16 Dec 2024): Mutual fund performance: the model for selecting persistent winners, The European Journal of Finance, DOI: 10.1080/1351847X.2024.2425401

To link to this article: https://doi.org/10.1080/1351847X.2024.2425401

THE EUROPEAN JOURNAL OF FINANCE
https://doi.org/10.1080/1351847X.2024.2425401

Routledge
Taylor & Francis Group

Mutual fund performance: the model for selecting persistent winners

Cesario Mateus[a], Irina B. Mateus[a] and Natasa Todorovic[b]

[a]Aalborg University Business School, Aalborg University, Aalborg, Denmark; [b]Bayes Business School, City, University of London, London, UK

ABSTRACT

Standard Fama-French-Carhart (FFC) models are widely used by academics to assess risk-adjusted fund performance versus market, size, style and momentum factors. However, they fail to reflect the industry standard, following which the performance of money managers is commonly evaluated relative to a corresponding benchmark and the peer group. In this paper, we introduce a new approach that augments the Carhart model and enables investors to identify the funds that outbid both the benchmark and the peer group. In addition, it allows discovering more certain winners by eliminating the under/outperformance of funds driven by the bias in the FFC factor construction. The application of our model is illustrated on Large Cap Value and Large Cap Growth US active equity mutual funds using contingency tables. The performance and persistence in performance are assessed by comparing our novel and the standard Carhart models. Our model identifies more winners than the Carhart; those winners earn higher returns net of benchmark and peer- group than the Carhart's winners, and show persistence in performance 36 months ahead. The results are robust to different specifications of contingency tables, holding periods or style categories of funds.

ARTICLE HISTORY
Received 3 July 2023
Accepted 24 September 2024

KEYWORDS
US equity mutual funds; Carhart model; benchmark-adjusted alphas; peer-group-adjusted alphas; performance persistence

JEL
G11; G12; G2

1. Introduction

Standard Fama and French (1993) and Carhart (1997) models are long-established models used as benchmarks in both asset pricing and mutual fund performance as well as persistence in performance literature. According to those models, mutual funds by and large do not create value for investors (see the review of literature on mutual fund performance and factor models in Cuthbertson, Nitzsche, and O'Sullivan (2010), and more recently Mateus et al. (2019a), Cremers, Fulkerson, and Riley (2019)). Even if it appears that active funds add value when gross returns are used to assess performance, the alphas are not statistically distinguishable from zero once the fees are accounted for and net returns are used to estimate performance (Berk and van Binsbergen 2015). Further, there is no conclusive evidence of long-term persistence in US mutual fund performance. If any persistence is documented, earlier studies in the area concluded that it is mostly confined to loser funds (e.g. Cai, Cheng, and Cheng 2018; Carhart 1997; Teo and Woo 2001).[1] More recently, Choi and Zhao (2021) find that any performance from Carhart (1997) disappears in recent years. Cuthbertson, Nitzsche, and O'Sullivan (2022) document that persistence can be identified only under specific circumstances: when investors use bootstrapping technique to select small number of top funds and hold them over short period of time. Similar was found in Cuthbertson, Nitzsche, and O'Sullivan (2023). However, while the standard models show whether funds have out/underperformed or persisted in performance given the embedded factor risks – the market, size, investment style and momentum – they do not provide a direct indication that is of high importance to investors and practitioners. In particular, they do not address what is funds' relative performance with respect to its peers and self-reported benchmark index – can funds outbid both their peer-group and their benchmark? If so, can such

CONTACT Cesario Mateus cmateus@business.aau.dk

Supplemental data for this article can be accessed online at https://doi.org/10.1080/1351847X.2024.2425401

performance persist? In this paper, we aim to answer these questions. Specifically, we introduce a new model that is of significant interest to the investment management industry: a comprehensive model that enables investors to identify mutual funds that outperform both their benchmark *and* the peer group. We demonstrate that those winner funds exhibit significant persistence in performance.

A recent strand of factor model literature identifies a bias in standard Fama-French and Carhart models (hereafter FFC models) in that they produce non-zero alphas when applied to passive indices typically used as mutual fund benchmarks (Chan, Dimmock, and Lakonishok 2009; Chinthalapati, Mateus, and Todorovic 2017; Cremers, Petajisto, and Zitzewitz 2012; Matallin-Saez 2007). Given that funds' stock selection is driven by the composition of their benchmarks, a fund's actual performance in the FFC models will be under(over)-stated during periods of benchmark underperformance (outperformance).[2] This highlights a flaw in the standard FFC models when it comes to identifying the stock-picking ability of mutual fund managers and/or future winners, and questions some of the existing evidence on performance and performance persistence. Angelidis, Giamouridis, and Tessaromatis (2013) and Chinthalapati, Mateus, and Todorovic (2017) develop models that account for benchmark performance, albeit not accounting for the performance of the peer group. Both studies find a significant improvement in the US equity fund alphas in the overlapping periods of their sample (1998-2012) following benchmark adjustment, indicating that throughout the period of their analysis benchmark indices produced overall negative FFC alphas. This implies that during periods of benchmark underperformance standard FFC models can wrongly classify a fund as a loser fund and vice versa, which we illustrate with example in Section 3.1 of the paper. Similar is documented for the UK market in Mateus, Mateus, and Todorovic (2016).

Note that a fund can outperform its benchmark yet be a poor performer within its peer group. Further, funds within a peer group have common idiosyncratic risks that may not be accounted for in the standard FFC or benchmark-adjusted models we've just discussed. Hunter et al. (2014) develop Active Peer Benchmark (APB) as an equally weighted portfolio of all the funds within a peer group, in an attempt to mitigate this issue. The APB is used to create an adjustment factor that modifies the standard Carhart model and produces peer-group adjusted alpha. Hunter et al. (2014) show that selecting winner funds based on this model will be a more successful strategy one year ahead than selecting winners via the Carhart method for US equity and bond mutual funds. Similar is documented in Mateus, Mateus, and Todorovic (2019b), who apply Hunter et al. (2014) model to the UK equity funds. However, while this method has success in identifying the funds with the greatest skill once accounted for the peer group, it does not tell us if the fund in fact outperformed its prospectus benchmark. It is possible that the whole peer group did worse than a passive benchmark index commonly used as a benchmark for that group. Therefore, selecting funds based on both benchmark-adjusted *and* peer-group adjusted performance can warrant better future performance than using either of these models separately (Mateus, Mateus, and Todorovic 2023).

Further, popular and commonly reported measures of performance such as FFC alphas, which are the focus of this paper, are not designed to provide direct insight into performance of a fund relative to their prospectus benchmark or other funds in the peer-group. To make the comparison, and select the funds that outbid the benchmark and the peer-group using standard measures of performance, one would have to compute the same measures for the fund, prospectus benchmark, and the peer-group and then determine whether the difference in their performance is statistically significant. Our model provides a framework in which a fund's performance relative to the self-designated benchmark *and* the peer-group is reflected in a single parameter: the benchmark-and-peer group adjusted alpha, as described in Section 2.2.

Our main contribution in this paper is methodological. To the best of our knowledge, this paper is the first to suggest an unbiased model that can identify the funds that outperform both their benchmark *and* their peer group simultaneously and show persistence in performance. This novel approach is of significant importance to investors and practitioners as it provides insights into the *relative* fund performance assessment required by the industry. Our model contributes to the literature on mutual fund performance measurement and performance persistence.

To illustrate our method, using the Morningstar category classification,[3] we select the funds in two US equity mutual fund categories – Large Cap Value and Large Cap Growth – and treat the categories as our peer groups. Our sample comprises 603 Large Cap Value and 785 Large Cap Growth active long-only equity mutual funds. We

use Morningstar to obtain information on the fund's primary prospectus benchmarks. Our method has three steps. First, in the spirit of Hunter et al. (2014), we compute Active Peer Benchmarks (APBs) as equally weighted portfolios of funds in each of our peer-groups: Large-Cap Value and Large Cap Growth. Second, in spirit of Angelidis, Giamouridis, and Tessaromatis (2013), we adjust the left-hand side of the Carhart (1997) model to obtain return of APB adjusted for the return of the benchmark corresponding to the peer-group, thus estimating benchmark-adjusted alpha of the APB. Third, we estimate benchmark- and peer group-adjusted alpha for each fund by augmenting the standard Carhart model by fund's benchmark adjusted return on the left-hand side and adding peer-group adjustment factor (benchmark adjusted alpha of APB plus the residual from the second step) on the right-hand side of Carhart equation. We show that mathematically, fund's Carhart alpha is equivalent to the sum of our new, augmented alpha, alpha of the fund's self-declared benchmark and the product of fund's sensitivity to the peer-group with the peer-group's benchmark adjusted alpha. We show that our model is superior in selecting winners to the standard Carhart (1997) model: (1) it allows for identifying more certain winners by eliminating the under(out)-performance of funds driven by the bias in the FFC factor construction (Cremers, Petajisto, and Zitzewitz 2012; Elton and Gruber 2020), (2) it identifies the best performers relative to the selected benchmark *and* the peer-group (formalising the empirical approach from Mateus, Mateus, and Todorovic 2023), (3) the winners selected by our model exhibit stronger significant persistence in performance 36 months ahead when compared with the winners suggested by the standard Carhart (1997) model.

We find that our augmented model selects more winners compared to Carhart model, which identifies more loser funds. While there is evidence of persistence in contingency tables based on both our new and the Carhart model, it is driven by winners in our model and by losers in the Carhart model. The winners from our model consistently earn greater excess returns net of benchmark and the peer-group than the Carhart's winners. Similarly, losers from our augmented model have more pronounced negative excess returns. This evidence indicates that our model is more effective in identifying both the 'true' winners and 'losers', outperforming both their benchmark and the peer-group. We conduct a number of robustness tests and demonstrate our model's ability to select persistent winners under alternative contingency tables specifications, holding periods and fund categories (peer-groups).

The rest of the paper is organised as follows. Section 2 presents the data and our new methodology. Section 3 lays out the results and Section 4 provides conclusions.

2. Data and methodology

2.1. Data

Our data is collected from Morningstar Direct. Our methodology can be applied to *any* peer-group an investor is interested in, and its output will be a selection of funds within that peer-group that outperform both the peer-group and fund's self-designated benchmark. Therefore, to illustrate our methodology, we focus on US active equity funds from two of the largest Morningstar style categories.[4] Specifically, after screening out index funds, and imposing a minimum requirement of 36 months of returns for each fund, our sample comprises of 603 active US long-only equity mutual funds that Morningstar places in the Large Cap Value category and 785 funds in the Large Cap Growth category, used as peer-groups in this study. Our sample is free of survivorship bias, including both 'alive' and 'dead' funds. For each fund, the oldest share class is selected, in line with many mutual fund studies, such as Cuthbertson, Nitzsche, and O'Sullivan (2022, 2023) or Clare et al. (2021) more recently. Our data spans from January 1990 to December 2020. For each fund, the monthly returns[5] net of fees and inclusive of dividends are collected. Morningstar also provides us with information on the fund's primary prospectus benchmark and benchmark's total returns. Table 1 provides information on some additional characteristics of funds in our two peer-groups. Large Growth funds are on average larger than Large Value funds, have higher net expense ratios and slightly longer manager tenure. Funds in both peer-groups spend on average around 14.5 years in our sample. Large Value funds report 24 different – prospectus benchmarks, while the equivalent number for Large Growth funds is 34. Developing the model that identifies the funds that can successfully outbid these respective benchmarks and their peer-group is the central focus of this paper.

Table 1. Description of fund characteristics in Large Cap Value and Large Cap Growth categories.

	Large Cap Value	Large Cap Growth
Number of unique funds	603	785
Average number of months in sample	175	178
Number of self-reported primary benchmarks within the category	24	34
Fund Size (in million USD, average)	4061	7234
Net Expense Ratio (Average)	0.94%	1.02%
Manager Tenure (average, in years)	9.42	10.03

Notes: The table reports the number of unique funds, average number of months in the sample, average fund size (in million od USD), average net expense ratio and average manager tenure (in years) for the Large Cap Value and Large Cap growth peer group.

2.2. The new model

In practice, mutual funds commonly report performance relative to their self-selected prospectus benchmark and relative to the peer group. On the other hand, the standard Carhart model is commonly used in fund performance evaluation, however, it examines a fund's performance in absolute terms, estimating whether a fund generates positive or negative excess performance given the four risk factors: market, size, style and momentum. Positive (negative) Carhart alpha funds are considered in the mutual fund performance literature as those that add (loose) value to investors. However, it is a stylised fact for mutual funds to report their performance relative to their prospectus benchmark and relative to the peer group (e.g. in a fund's fact sheet). Similarly, investors are interested in selecting the top performing funds within a category (peer-group) of their choice (say, large-cap growth) that have performed better than their benchmark. It is therefore important to emphasise that the existing mutual fund performance studies based on Carhart alpha reveal nothing about the fund's performance *relative* to its self-declared prospectus benchmark and the peer group. In fact, Carhart alpha may classify a fund as an underperformer when a fund outperforms the benchmark and all other funds in the peer group.[6] Existing modifications of the Carhart model either account for the benchmark (Angelidis, Giamouridis, and Tessaromatis 2013; Chinthalapati, Mateus, and Todorovic 2017) or the peer group (Hunter et al. 2014; Mateus, Mateus, and Todorovic 2019b), but not both. Our new model, outlined below, will enable investors to select funds that outperform other similar funds on the market *and* the given benchmark index at the same time.

Our methodology involves modification of both the left and the right-hand side of the standard Carhart (1997) model. The standard Carhart model, used as a baseline in our study, is specified as:

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

where $R_{i,t} - R_{f,t}$ is the excess return of fund i in period t, and α_i is the fund's four-factor Carhart alpha, signifying a fund's value-added given four risk factors. $(R_{M,t} - R_{f,t})$ is the market risk premium, defined as the excess return of the broad US market portfolio, represented by a value-weighted portfolio of all NYSE, AMEX and NASDAQ listed firms, in excess of one-month T-bill rate. Please note that this 'market portfolio' is broader portfolio than any of the indices used as benchmarks in this study, so any benchmark index used in this paper is only a sub-set of the 'market'. As per Fama and French (1993), SMB is a proxy for the size risk representing the difference in returns between the small and the big cap stock portfolio; HML is the proxy for value risk, and it is obtained as the difference between the returns on a portfolio of stocks with high-book-to-market ratio and a portfolio of low-book-to-market ratio stocks. Finally, as per Carhart (1997), WML is the momentum factor defined as the difference between returns on a portfolio of past winner stocks minus the returns on a portfolio of past loser stocks. All factors and their definitions are standard, as in Carhart (1997), and obtained from Kenneth French's data library.[7]

To account for the factors common to the peer group but not accounted for in the Carhart model, in the spirit of Hunter et al. (2014), we first construct the Active Peer Benchmark as an equally weighted return of all the funds within the peer-group (APB). We then examine whether the peer-group as a whole (APB) produces excess performance relative to the investment style it represents, i.e. we isolate commonalities of the peer group are not embedded in the benchmark index representative of the investment style of the peer group. We achieve this by estimating the alpha of APB using the augmented version of the Carhart model where the risk-free

rate is substituted with the return of the benchmark on the left-hand side, as in Angelidis, Giamouridis, and Tessaromatis (2013) and Mateus et al. (2019b):

$$R_{APB,t} - R_{b,t} = \alpha_{APB,b} + \beta_{APB,b,M}(R_{M,t} - R_{f,t}) + \beta_{APB,b,SMB}SMB_t + \beta_{APB,b,HML}HML_t + \beta_{APB,b,WML}WML_t + e_{APB,b,t} \quad (2)$$

where $R_{APB,t}$ is the return of APB portfolio of each of the peer group we examine, $R_{b,t}$ is the return of an index used as a representative benchmark for the investment style of our peer-group (in our case, Russell 1000 Value for Large Cap Value and Russell 1000 Growth for Large Cap Growth); $\alpha_{APB,b}$ is the benchmark-adjusted alpha of the peer-group and $e_{APB,b,t}$ is the difference of the idiosyncratic risk (error terms) of the peer-group and of the benchmark index with respect to the common four factors in the model. If $\beta_{APB,b,M}, \beta_{APB,b,SMB}, \beta_{APB,b,HML}$ and $\beta_{APB,b,WML}$ are different from zero – the APB portfolio has the factor loadings different from those of the corresponding benchmark, e.g. if $\beta_{APB,b,SMB} = 0.05$ implies that the peer group has 5% more small-cap stocks than the benchmark index representative of the investment style of the peer-group.

Finally, in our new model presented in Equation (3), we modify both the left and the right hand-side of the Carhart's model (Equation (1)), to obtain the benchmark-and-peer-group adjusted performance of a mutual fund. Specifically, on the right hand-side of our model we incorporate the unique performance and risk of the fund's peer-group, by adding the 'peer-group factor' ($\alpha_{APB,b} + e_{APB,b,t}$) obtained from Equation (2). On the left hand-side of Equation (3), we account for the funds self-declared benchmark, by replacing the risk-free rate from Equation (1) with the return of the fund's self-declared primary prospectus benchmark. The alpha (intercept) resulting from this Equation (3) is the measure of benchmark – and peer-group adjusted-performance of the fund. Our new, augmented model is therefore given as:

$$R_{i,t} - R_{b,t} = \alpha_{i,AGM} + \beta_{ib,AGM,M}(R_{M,t} - R_{f,t}) + \beta_{ib,SMB}SMB_t + \beta_{ib,HML}HML_t + \beta_{ib,WML}WML_t + \beta_{ib,APB}(\alpha_{APB,b} + e_{APB,b,t}) + \omega_{i,AGM,t} \quad (3)$$

where $R_{i,t} - R_{bi,t}$ represents the benchmark adjusted return of fund i, obtained using fund's self-reported primary prospectus benchmark[8]; $(\alpha_{APB,b} + e_{APB,b,t})$is a factor used to amend the right-hand side of the Carhart model, which represents commonality in performance and risk unique to the peer group, not embedded in the peer-group's style benchmark, from Equation (2); $\alpha_{i,AGM}$ is the benchmark and peer-group augmented alpha of fund i; $\omega_{i,AGM,t}$ is representing the difference between unique risks of the fund and its prospectus benchmark with respect to the five factors in this model; $\beta_{ib,APB}$ is the fund's sensitivity to the peer-group factor $(\alpha_{APB,b} + e_{APB,b,t})$, beyond the benchmark; and $\beta_{ib,M}, \beta_{ib,SMB}, \beta_{ib,HML}, \beta_{ib,WML}$ are fund's market, size, value and momentum factor loadings in excess of its self-reported benchmark.

To highlight the relationship between a fund's Carhart alpha and augmented benchmark-and-peer-group-adjusted alpha from our new model from Equation (3), we will present an alternative specification for this approach. The equivalent of the model from Equation (3) is given as:

$$R_{i,t} - R_{b,t} = \alpha_{i,AGM} + \beta_{ib,APB}(R_{APB,t} - R_{b,t}) + \beta_{ib,M}(R_{M,t} - R_{f,t}) + \beta_{ib,SMB}SMB_t + \beta_{ib,HML}HML_t + \beta_{ib,WML}WML_t + \omega_{i,AGM,t} \quad (4)$$

Following Frisch–Waugh–Lowell theorem, the model from Equation (4) can be expressed as the regression of residuals of benchmark-adjusted return of a fund $(R_{i,t} - R_{b,t})$ on Carhart's excess return of the market, SMB, HML and WML risk factors[9] on the residuals obtained by regressing active peer group portfolio return adjusted for the peer-group's benchmark $(R_{APB,t} - R_{b,t})$ on Carhart's four factors (as in Equation (2)):

$$e_{ib,t} = \alpha_{i,AGM} + \beta_{ib,APB}e_{APB,b,t} + \omega_{i,b,t} \quad (5)$$

Further, using de-meaned residuals $\tilde{e}_{APB,b,t}$, of the regression in Equation (2), we can express $e_{ib,t}$ as:

$$e_{ib,t} = \alpha_{ib} + \beta_{ib,APB}\tilde{e}_{APB,b,t} + \omega_{i,b,t} \quad (6)$$

The intercept of Equation (6), fund's benchmark-adjusted alpha, α_{ib}, satisfies the following relationship:

$$\alpha_{i,b} = \alpha_{i,AGM} + \beta_{ib,APB}\alpha_{APB,b} \tag{7}$$

Further, since benchmark-adjusted alpha of a fund, α_{ib}, represents the difference between fund's and benchmark' Carhart alpha ($\alpha_{i,b} = \alpha_{i,Carhart} - \alpha_{benchmark,Carhart}$), then, our augmented, benchmark and peer-group adjusted alpha can be expressed as:

$$\alpha_{i,AGM} = \alpha_{i,Carhart} - \alpha_{benchmark,Carhart} - \beta_{ib,APB}\alpha_{APB,b} \tag{8}$$

Therefore, our augmented alpha from model specification in Equation (4) represents excess return conditional on Carhart's risk factors, alpha of the fund's prospectus benchmark and local peer-group risk factor given by ($R_{APB,t} - R_{b,t}$). Simple re-arranging of Equation (8) highlights the relationship between the standard Carhart and our new model alpha, and brings to light the bias embedded in fund's Carhart alpha:

$$\alpha_{i,Carhart} = \alpha_{i,AGM} + \alpha_{benchmark,Carhart} + \beta_{ib,APB}\alpha_{APB,b} \tag{9}$$

The equation shows that fund's Carhart's alpha is a linear function of the benchmark's Carhart alpha, fund's sensitivity to the peer-group, $\beta_{ib,APB}$, and the peer-group's benchmark-adjusted alpha. This implies that the more fund is related to the peer-group and the higher the alpha of both the peer-group and the benchmark, the more bias there is in the fund's Carhart alpha. When the passive benchmark index used as a fund's benchmark and its peer group do not generate alpha, our augmented model will yield the same alpha as the standard Carhart model.

Note that the model we present is a relative performance model, dependent on the choice of benchmark (as in Angelidis, Giamouridis, and Tessaromatis 2013; Chinthalapati, Mateus, and Todorovic 2017; Mateus, Mateus, and Todorovic 2016, 2023) and the peer-group (as in Hunter et al. 2014 or Mateus, Mateus, and Todorovic 2019c). Regarding the peer-group, we believe that investors investment preferences and risk appetite will determine the peer-group of funds they would like to invest in (within which they would want to make a performance comparison using our model). As far as choice of benchmark is concerned, funds commonly report their performance alongside their self-designated prospectus benchmark. It is very rare for a fund to change the benchmark, so choosing a fund's primary prospectus benchmark to gouge their benchmark-adjusted segment of performance would be appropriate in our model.

3. Empirical results

3.1. Preliminary results

To start, it is worth noting that if a performance measure is appropriately specified, it should not result in out/under performance when applied to passive benchmarks (Chen and Knez 1996). We provide an illustration that this is not the case when it comes to Carhart's measure. There is evidence of presence of non-zero Carhart alphas for the typical benchmarks in Large Cap Value and Large Cap Growth categories of funds: Russell 1000 Value and Russell 1000 Growth index. Note that in the Carhart model, the market risk premium is based on a broad value-weighted market portfolio from a cross section of NYSE, AMEX and Nasdaq Stocks and unlike Russell 1000 Value (or Growth) Index, not focusing on any particular segment of the US market. Therefore, if one estimates the Carhart model (Equation (1)) for a benchmark index (i.e. the excess return of a fund is replaced with excess return of a benchmark on the left-hand side of Equation (1)), one would obtain the alpha of the benchmark index (in this case Russell 1000 Value or Russell 1000 Growth). Existing empirical evidence is in support of the fact that Carhart alphas of various passive indices are often different from zero (e.g. Cremers, Petajisto, and Zitzewitz 2012, Mateus et al., 2016). Therefore, a fund could claim to be active, yet simply engage in mimicking their benchmark index: if that benchmark generates significantly positive Carhart alpha – so will the fund. Such alpha, therefore, does not represent fund manager's 'true' stock-picking ability. Our new model will enable investors to avoid such funds as well as funds whose positive excess returns are driven by mimicking a strategy of an outperforming peer-group, not by their unique skill.

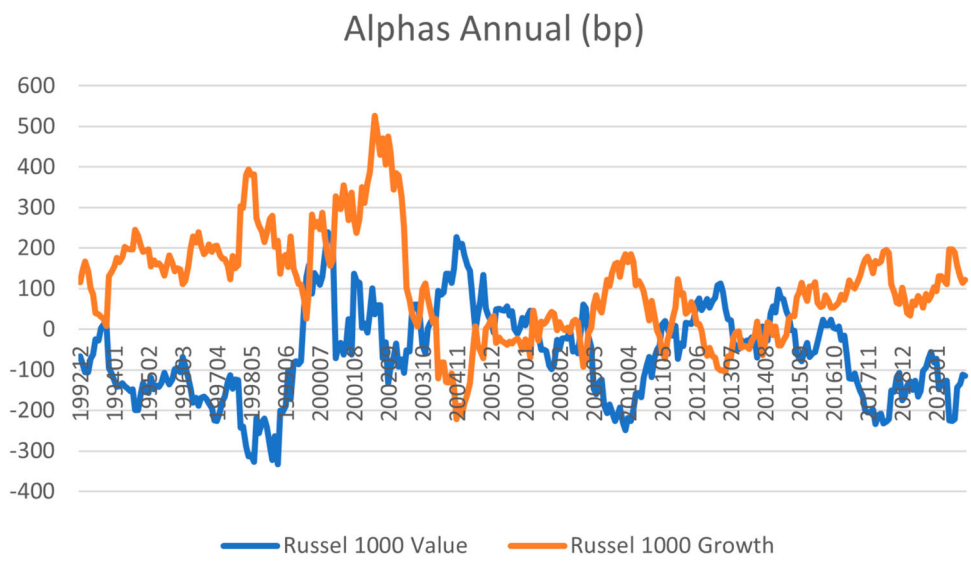

Figure 1. Rolling Carhart alpha of Russell 1000 Value and Russell 1000 Growth Index.
Notes: The chart represents Carhart alphas of Russell 1000 index January 1990 to December 2020, estimated over 336 rolling periods, each being 36 months long and rolled over by 1 month: Jan 1990–Dec 1992, Feb 1990–Jan 1993 and so on until the end of sample period is reached, Jan 2018–Dec 2020.

Figure 1 illustrates that both Russell 1000 Value index and Russell 1000 Growth index exhibit periods of very large non-zero alphas in our sample. Carhart alphas for each index are estimated over 336 rolling windows (each being 36 months in length), rolled forward by 1 month: January 1990–December 1992, February 1990–January 1993 and so on. The two indices exhibit the opposite trends, in line with what one would expect from portfolios with holdings at different end of spectrum in terms of P/E ratios and dividend yields. During the height of the dot.com boom in early 2000s, Russell 1000 growth index exhibits very large positive alphas in excess of 500 bp p.a. Following the burst of the dot.com bubble the alpha of Russell 1000 Growth falls to −200bps. Russell 1000 Value has less pronounced peaks and throughs of alphas but nevertheless exhibits large negative alphas in both the 1990s and following financial crisis while it generated positive performance during dot.com bubble burst. The significance of these alphas (t-stats) also varies over time; stronger significance is observed in the known boom/crises periods in our sample (the dot.com boom/burst or the financial crisis).[10] Such high and low alphas embedded in the passive indices, commonly used as benchmarks in the Large Cap Value and Large Cap Growth peer groups of our funds are found to impact the accuracy of the standard Carhart model in estimating mutual fund performance and performance persistence as demonstrated in our methodology.

According to Mateus, Mateus, and Todorovic (2016), the presence of negative alpha in a passive index used in their study amplifies fund underperformance, i.e. once adjusted, fund's Carhart alphas are found to be lower than their benchmark-adjusted four-factor alphas. In addition, Carhart model can report a negative alpha for a fund that has performed better than their peer-group, even though that is a trait desirable for investors.

To illustrate this point that Carhart model can wrongly classify a fund as underperformer, particularly during benchmark (and peer-group) underperformance, we present an example of one fund from our sample, we call it a Growth Fund. The illustration is presented in Table 2.

Table 2 shows that following a 36-month estimation period within our sample, during which benchmark index performance is negative (Carhart alpha of −63.09 bp p.a.), our Growth Fund is classified as underperformer by Carhart's alpha (−61.74 bp p.a., significant at 1%). If an investor followed fund selection based on Carhart's alpha, they would not have selected this fund although it has actually performed better than both their benchmark (Russell 1000 Growth index) and the peer-group during the estimation period. Should the investor follow our new model, which produces augmented alpha of 284.69p.a. at the time of index underperformance in the estimation period, they would have generated 1.70% higher return than the benchmark and

Table 2. Illustration of growth fund's new and Carhart alphas during benchmark index underperformance.

_	Estimation period $t-36$ m	Performance period $t+36$ m
Our new model alpha p.a.	284.69 bp, significant at 1%	358.93 bp, significant at 1%
Carhart alpha p.a.	−61.74 bp, significant at 1%	229.54 bp, not significant
Benchmark (Russell 1000 Growth) Carhart alpha p.a.	−63.09 bp	114.42bp
Peer-group (APB) Carhart alpha p.a.	−229.70 bp	−131.169bp
Excess return benchmark p.a.	—	18.07%
Excess return peer-group p.a.	—	16.51%
Excess return Growth Fund p.a.	—	19.77%
Excess return of Growth Fund above the benchmark and the peer-group p.a.	—	1.70%

Notes: The table presents the alpha from the new model and the Carhart model for Growth Fund estimated over 36-month period, the Carhart alpha of the benchmark index (Russell 1000 Growth) and the peer group over the same 36-month estimation period. The table also shows post-estimation performance in the next 36 months, including Growth Fund's augmented new alpha, Carhart alpha, and excess returns as well as benchmark's and peer-groups Carhart alpha and excess returns in the same post-estimation period. All reported values are annualised.

Table 3. Number of positive and negative alphas (new and Carhart models) at the time of benchmark index under/out performance.

Panel A: Large Cap Value Style			
When Russell 1000 Value alpha is negative …		When Russel 1000 Value alpha is positive …	
# Positive alphas new model	# Positive alphas, Carhart	# Positive alphas new model	# Positive alphas, Carhart
30,669	16,462	26,842	19,447
# Negative alphas new model	# Negative alphas, Carhart	# Negative alphas new model	# Negative alphas, Carhart
26,804	41,011	26,373	33,768
Panel B: Large Cap Growth Style			
When Russell 1000 Growth alpha is negative …		When Russel 1000 Growth alpha is positive …	
# Positive alphas new model	# Positive alphas Carhart	# Positive alphas new model	# Positive alphas, Carhart
15,137	7872	40,349	36,537
# Negative alphas new model	# Negative alphas Carhart	# Negative alphas new model	# Negative alphas, Carhart
16,753	24,018	45,360	49,172

Notes: The table shows number of positive and negative alphas at the time of Russell 1000 Value benchmark index under/out performance in Panel A and Russell 1000 Growth index in panel B. Alphas for Large Cap Value funds (Panel A) and Large Cap Growth funds (Panel B) are estimated using 36-month rolling periods starting from January 1990, rolled forward by 1 month until December 2020, 336 months in total.

the peer-group in the post-estimation period of the next 36 months, the new model alpha of 358.93 bp p.a. and higher Carhart alpha than both the benchmark and the peer-group. It is also interesting to note that this Growth Fund would generate economically (albeit not statistically) significant Carhart alpha of 229.54 bp in the next 36-months. Therefore, this illustration shows that our model helps correct biases in Carhart alpha that wrongly classify some funds as losers, even though they perform better than both the benchmark and the peer-group, which is of interest to investors.

As a further preliminary step to our analysis, we evaluate the impact of the bias caused by benchmark's non-zero Carhart alphas and the impact of the peer-group on our sample of funds. To illustrate the impact of the two biases embedded in the standard Carhart factor model in our sample, we separate our sample into 36-monthly periods of negative and positive performance (Carhart alpha) of the benchmark index most representative of the peer group, starting from January 1990, rolled forward on a monthly basis until we reach the end of our sample. We then count the number of positive and negative fund alphas generated by our funds based on (1) our new augmented model and (2) the standard Carhart model in each of those periods. Table 3 presents the summary of these preliminary results. Panel A of Table 3 shows the number of positive and negative alphas in the Large Cap Value category of funds when Russell 1000 Value index alpha is positive or negative and Panel B shows the same for Large Cap Growth category of funds and Russell 1000 Growth index.

Panel A points that both at times of index out- and underperformance, the Carhart model ascertains a greater number of fund-periods with negative alphas than positive alphas (41,011 vs. 16,462 when index alpha is negative and 33,768 vs 19,447 when index alpha is positive). Moreover, it identifies a greater number of negative alpha funds compared to our new model (e.g. 41,011 Carhart vs. 26,804 new model negative alphas in the periods of index underperformance). Panel B shows the same pattern in ratio between negative and positive Carhart alphas

in both index states, where the number of negative alphas dominates the number of positive ones, particularly at times of underperformance of Russell 1000 Growth index with respect to the four factors. By comparison, in both Panel A and Panel B, our new model identifies more positive and less negative alphas than the Carhart model. This implies that a fund that underperforms when Carhart's four risk factors are considered may still be better than both their benchmark and other funds in the peer group. This is particularly pronounced in periods of poor index performance, where the number of positive alphas from our model we examine is nearly double than the number of positive Carhart's alphas in both Panel A (Large Cap Value funds) and Panel B (Large Cap Growth funds). This line with Mateus, Mateus, and Todorovic (2016), who state that in the periods of benchmark index underperformance which dominate their study, the benchmark-adjusted alphas are higher than the Carhart alphas. In the periods of benchmark index outperformance, our model continues to produce greater number of positive alphas relative to the Carhart model.

Overall, these preliminary results highlight potential deficiency of Carhart alpha in identifying true winners that can outperform their benchmark and their peer-group. Our proposed new model corrects for these biases. To formally demonstrate the advantage of our new benchmark and peer-group adjusted model relative to the standard Carhart model in selecting the 'true winner' funds (outperforming both the benchmark and the peer-group), we construct contingency tables and test performance persistence, as described in the next section.

3.2. Contingency tables and performance persistence

We construct standard contingency tables following Mateus et al. (2019b, 2023) and Fletcher and Forbes (2002), among many others. The contingency tables are based on four groups of funds (2 × 2 tables): winner/winner (W/W), winner/loser (W/L), loser/winner (L/W) and loser/loser (L/L). To define winners and losers, we estimate funds' alphas using 36 months[11] of non-overlapping returns over two consecutive periods: α_t (based on $t-36$ m of returns) and α_{t+1} (estimated using $t+36$ m of returns) rolled over by one month until we reach the end of our sample. Specifically, the first α_t for each fund is based on January 1990–December 1992 ($t-36$ m) returns, α_{t+1} is based on January 1993–December 1995 returns ($t+36$ m); we then roll estimation period by one month to estimate the next pair of α_t and α_{t+1}, using returns February 1990–January 1993 for α_t and February 1993–January 1996 for α_{t+1}; and so on. In total, there are 300 rolling periods.

We consider statistical significance of alphas for the definition of winner and loser categories in our contingency tables, and gouge significance of alphas using a standard t-test. Hence, we define W/W as those funds with $\alpha_t > 0$, statistically significant at 10% level (t-stat > 1.65), and $\alpha_{t+1} > 0$ (both significant and insignificant).[12] L/L funds are those with significant $\alpha_t < 0$ (at 10%), followed by $\alpha_{t+1} < 0$ (both significant and insignificant). The intermediate groups, winner/loser (W/L) and loser/winner (L/W) are comprised of: funds that have significant $\alpha_t > 0$ and significant or insignificant $\alpha_{t+1} < 0$ (W/L); and funds with significant $\alpha_t < 0$ and significant or insignificant $\alpha_{t+1} > 0$ (L/W). We construct two contingency tables (CTs hereafter) based on (1) alphas from our new augmented model ($\alpha_{i,AGM}$ from Equation (3)) and (2) standard Carhart alphas (α_i from Equation (1)), for comparison.

Table 4 lays out the results for both CTs, based on our new augmented model and Carhart model. Panel A reports results for Large Cap Value and Panel B for Large Cap Growth category of funds. In both panels, the CT based on our new augmented alphas reports over two times more winners (8437 in Panel A and 8748 in Panel B) than the Carhart model (3944 in Panel A and 4061 in Panel B), while Carhart identifies 1.5–2 times more funds as losers (13,189 in Panel A; 11,802 in Panel B) compared to our augmented model (6304 in Panel A and 8051 in Panel B). Significant persistence in performance is documented for funds classified by the Carhart model, but it is driven by loser funds[13]: there is six times more L/L funds (9803) than W/W funds (1629) in Panel A. This result is consistent with a number of performance persistence studies in mutual fund literature who state that persistence is driven by loser funds (see Carhart 1997). In contrast, using our new augmented model to classify the funds, we find that W/W funds are the drivers of significant persistence in performance (χ^2 of the CT is significant at 1%). We identify 5070 Large Cap Value (Panel A) and 5562 Large Cap Growth W/W funds (Panel B) – a number higher than in any other group in the CT table for our augmented model. To gauge the economic significance of these results for investors, we calculate winner (and loser) funds' annualised excess returns above the benchmark and the peer group for period $t+36$ months, following investment in funds identified as winner

Table 4. Contingency tables augmented (new) and Carhart models.

Panel A: Large Cap Value Funds

Contingency table, Augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	5070	3367	8437	Winner	1629	2315	3944
Loser	2842	3462	6304	Loser	3386	9803	13, 189
Total	7912	6829	14, 741	Total	5015	12, 118	17, 133

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model
Post Estimation Period: $t + 36$ m

Winner funds	+56.04 bp p.a.
Loser funds	−43.83 bp p.a.

Panel B: Large Cap Growth Funds

Contingency table, Augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	5562	3186	8748	Winner	1630	2431	4061
Loser	2642	5409	8051	Loser	3820	7982	11,802
Total	8204	8495	16,799	Total	5450	10,413	15,863

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model
Post Estimation Period: $t + 36$ m

Winner funds	77.65 bp p.a.
Loser funds	−39.62 bp p.a.

Notes: The upper part of Panel A (Large Cap Value funds) and Panel B (Large Cap Growth funds) of the table reports 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t - 36$ m and positive (negative) alphas in period $t + 36$ m (rolled one month forward). Lower part of the table reports returns (in bp p.a.) the difference among the augmented and the Carhart models in excess of the benchmark and the peer-group in period $t + 36$ m for all winner and loser funds.

funds and loser funds based on their $t - 36$ m alphas estimated using (i) our new model and (ii) Carhart. The lower part of each panel of Table 4 reports the difference in those excess returns between winner (and loser) funds selected by our augmented model and those selected by Carhart model. We find that Large Cap Value winners from our augmented model earn benchmark and peer-group adjusted return which is 56.05 bp above that of the Carhart winners (Large Cap Growth winners gain 77.65 bp over winner funds selected by the Carhart model). Similarly, loser funds from the Large Cap Value (Large Cap Growth) category selected by our augmented model have more pronounced negative excess returns, generating −43.83 bp p.a. (−39.62 bp p.a.) lower excess returns than losers from the Carhart model. This corroborates our earlier notion that our new, augmented model is more successful in identifying both the 'true' winners and 'losers'.

3.2.1. Unique winners and losers

To highlight the difference in the ability of the two models to select winners (and losers), we construct a standard 2 × 2 contingency table (as Table 4) for funds that are (a) uniquely identified as winners or losers delete in our augmented model (but not in Carhart) and (b) uniquely identified as winners and losers in Carhart (but not in our augmented) model. Table 5 presents the results. Panel A for Large Cap Value funds reports that Carhart model identifies 574 unique funds that are not picked up by our augmented model that start as winners in the estimation period $t - 36$ m, and only 155 of those are in the W/W category. Our model on the other hand identifies 5067 unique winners that outperform the peer group and the benchmark in the estimation period, not picked up by the Carhart model. Out of those, 63% are W/W funds. On the other hand, Carhart classifies nearly seven times more funds as losers compared to our augmented model. When it comes to persistence in performance, funds uniquely selected by our benchmark and peer-group adjusted model exhibit strong persistence (χ^2 significant at 1%), stemming from W/W (3177) rather than L/L (831) funds. Notably, the performance of

Table 5. Persistence in performance of unique winners and losers by (1) our augmented model and (2) by standard Carhart model.

Panel A: Large Cap Value Funds

Contingency table, New augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	3177	1890	5067	Winner	155	419	574
Loser	382	831	1213	Loser	1957	6141	8098
Total	3559	2721	6280	Total	2112	6560	8672

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model
Post Estimation Period: $t + 36$ m

Winner funds	+210.87 bp p.a.
Loser funds	−118.52 bp p.a.

Panel B: Large Cap Growth Funds

Contingency table, Augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	3787	2040	5,827	Winner	329	811	1140
Loser	687	1587	2274	Loser	2018	4007	6025
Total	4474	3627	8101	Total	2347	4818	7164

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model
Post Estimation Period: $t + 36$ m

Winner funds	+117.99 bp p.a.
Loser funds	−80.59 bp p.a.

Notes: Table reports 2 × 2 the same type of contingency table as Table 4 for funds that are uniquely selected as winners/losers by our augmented model but are not selected by Carhart model, and vice versa – funds that are uniquely selected as winners/losers based on Carhart model but not based on our augmented model. All else is as per notes in Table 4.

funds selected as winners or losers by the Carhart model (but not our model) does not persist (χ^2 has p-value of 0.13), proving that the Carhart model is inferior to our model when it comes to framing winners.

Reverting our attention to the difference in annualised returns in excess of peer-group and the benchmark in the period $t + 36$ months, unique funds selected as winners by our augmented model produce 210.87 bp higher excess return p.a. than those selected by Carhart, while losers from our model have −118.52 bp lower excess return p.a. than Carhart's losers.

In Panel B, the results for Large Cap growth funds are qualitatively the same as in Panel A, except that performance of funds uniquely selected by Carhart model also persists. Unlike for funds uniquely selected by our new model, the source of persistence in funds unique to Carhart is in loser funds: there are over 10 times more L/L funds than W/W funds (χ^2 significant at 1%).

These findings are echoing the point that our new model is more successful in selecting 'true' winner funds that will reap economic benefits for investors one period hence, and it will help investors identify more extreme future losers.

Note that we find no evidence that a particular choice of benchmark drives a fund to be classified as a winner or a loser. 88.56% of the funds in the overall Large Cap Value sample report one of the following three benchmarks: Russell 1000 Value (56%), S&P 500 (26%) and Russell 3000 Value (6%). Similarly, 84.33% of funds in the Large Cap Growth group report one of the following three benchmarks: Russell 1000 Growth (42%), S&P 500 (39%) and Russell 3000 Growth (3%). Very similar distribution across the three benchmarks is observed in both the winner and the loser group in each category of funds.

In the next section, we conduct a number of tests to demonstrate that results from Table 4 are robust to different specifications of measuring persistence in performance and apply to alternative investment horizons and fund categories.

Table 6. Contingency tables for augmented (new) and Carhart models: funds with significant alphas only.

Panel A: Large Cap Value Funds							
Contingency table, Augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	1051	274	1325	Winner	185	255	440
Loser	320	747	1067	Loser	150	2963	3113
Total	1371	1021	2392	Total	335	3218	3553
Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model Post Estimation Period: $t + 36$ m							
Winner funds	+56.04 bp						
Loser funds	−43.83 bp						
Panel B: Large Cap Growth Funds							
Contingency table, Augmented model				Contingency table, Carhart			
Augmented model	Winner	Loser	Total	Carhart	Winner	Loser	Total
Winner	1488	425	1913	Winner	129	600	729
Loser	297	1431	1728	Loser	369	2175	2544
Total	1856	1785	3641	Total	498	2775	3273
Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model Post Estimation Period: $t + 36$ m							
Winner funds	+77.65 bp						
Loser funds	−39.62 bp						

Notes: The upper part of Panel A (Large Cap Value funds) and Panel B (Large Cap Growth funds) of the table reports 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t - 36$ m and positive (negative) and significant alphas in period $t + 36$ m (rolled one month forward). Lower part of the table reports returns (in bp p.a.) the difference among the augmented and the Carhart models in excess of the benchmark and the peer-group in period $t + 36$ m for all winner and loser funds.

3.3. Robustness of the results

3.3.1. Including only statistically significant alphas

In the first robustness test, we repeat the approach used to construct Table 4, but we limit our fund universe only to those funds that are able to generate statistically significant alphas in both estimation period t (estimated using $t - 36$ m) and performance period $t + 1$ (estimated using $t + 36$ m). We define W/W funds as those that have $\alpha_t > 0$, and $\alpha_{t+1} > 0$, both statistically significant at 10% (based on a t-test from rolling regressions) and L/L funds have both $\alpha_t < 0$, and $\alpha_{t+1} < 0$ statistically significant at 10%. The two intermediate categories W/L, and L/W are combination winners and losers, where W/L (L/W) contains funds that have positive (negative) and significant alpha in period t and negative (positive) and significant alpha in period $t + 1$. Table 6 reports these results and corroborates our findings from Table 4. In both Panel A (Large Cap Value) and Panel B (Large Cap Growth) our new, benchmark and peer-group augmented alpha model identifies more winners than the Carhart model, and performance persistence (χ^2 significant at 1%) is driven by the winners (Large Cap Value) and both winners and losers (Large Cap Growth). In contrast, Carhart model identifies more losers than our model and persistence (χ^2 significant at 1%) is driven by the loser funds. Winner (loser) funds selected by our new model generate higher (lower) returns in excess of the benchmark and the peer group than winners (losers) from the Carhart model, in both panels of Table 6.

3.3.2. Including all the funds: 4 × 4 contingency table

Contingency tables in Tables 4–6 include only the funds that in estimation period t have alphas significant at 10% (the definition of winners and losers can be found in Section 3.2). To include all funds from our sample in the analysis, and to differentiate further the W/W and L/L categories, we construct a 4 × 4 contingency table. The 16 groups of fund alphas in consecutive periods t and $t + 1$ are determined as all possible combinations of

Table 7. 4 × 4 Contingency tables augmented (new) and Carhart models.

Panel A: Large Cap Value Funds

	Contingency table, Augmented model					Contingency table, Carhart					
Augmented model	Winner	Positive alpha	Negative alpha	Loser	Total	Carhart	Winner	Positive alpha	Negative alpha	Loser	Total
Winner	1051	4019	3093	274	8437	Winner	185	1444	2060	255	3944
Pos. alpha	3529	18,018	14,952	2662	39,161	Pos. alpha	915	8906	14,511	3771	28,103
Neg. alpha	2430	13,310	14,338	3129	33,207	Neg. alpha	875	11,236	21,832	7930	41,873
Loser	320	2522	2715	747	6304	Loser	150	3236	6840	2963	13,189
Total	7330	37,869	35,098	6812	87,109	Total	2125	24,822	45,243	14,919	87,109

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart
Post Estimation Period: $t + 36$ m

Winner funds	**Positive alpha funds**	**Negative alpha funds**	**Loser funds**
+56.04 bp	+20.17 bp	−21.33 bp	−43.83 bp

Panel B: Large Cap Growth Funds

	Contingency table, Augmented model					Contingency table, Carhart					
Augmented model	Winner	Positive alpha	Negative alpha	Loser	Total	Carhart	Winner	Positive alpha	Negative alpha	Loser	Total
Winner	1488	4074	2761	425	8748	Winner	129	1501	1831	600	4061
Pos. alpha	3692	14,686	14,887	2929	36,194	Pos. alpha	1199	10,379	15,321	3677	31,576
Neg. alpha	2142	13,407	17,585	4680	37,814	Neg. alpha	1538	14,148	21,792	5890	43,368
Loser	297	,345	,978	,431	8051	Loser	369	3451	5807	2175	11,802
Total	,619	34,512	39,211	9465	90,807	Total	3235	29,479	44,751	13,342	90,807

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart
Post Estimation Period: $t + 36$ m

Winner funds	**Positive alpha funds**	**Negative alpha funds**	**Loser funds**
+77.65 bp p.a.	+12.27 bp p.a.	−25.81 bp p.a.	−39.62 bp p.a.

the following groups: W/W are funds have positive alpha significant at the 10% level[14] in t and $t + 1$; 'Positive alpha funds' category have positive but insignificant alphas in t and $t + 1$; Negative alpha funds have alpha below zero albeit insignificant in t and $t + 1$; and, finally, L/L funds are those with negative and significant alphas (at 10% level) in both period t and $t + 1$. The 4 × 4 CT in Table 7 for both Large Cap Value (Panel A) and Large Cap Growth (Panel B) category confirms the results from all earlier tables, documenting significant persistence in performance (χ^2 significant at 1%) driven by winner and positive alpha funds selected by our new model.

Persistence in performance based on the Carhart model in Table 7 is driven by losers (χ^2 significant at 1%). Moreover, in the second part of the table in each panel we show that the difference in excess returns (beyond the benchmark and the peer-group) in period $t + 1$ between funds selected by our new model and those selected by Carhart model is decreasing across the four categories of funds. It is the highest between the winners from both models, gradually decreasing as we move to 'positive alpha' fund category, becoming negative when comparing excess returns of 'negative alpha' categories and, finally exhibiting the largest negative difference in loser funds group, confirming superior ability of our approach to select true winners (and losers).

3.3.3. Decile and quintile portfolio sorting

We test results achieved by our new model by constructing further two contingency tables based on recursive portfolio sorting where funds are split into deciles and quintiles according to their benchmark and peer group augmented alphas. Winners/Winners (Losers/Losers) are funds in the top (bottom) deciles and quintiles based on alphas in period t, (estimated using $t - 36$ months), and continue to be in the top (bottom) deciles and quintiles based on alphas in period $t + 36$ m (rolled one month forward, over 300 rolling periods). The top part

Table 8. Contingency tables augmented (new) and Carhart models (Deciles and Quintiles).

Panel A: Large Cap Value							
Contingency table, Augmented model (Deciles)				Contingency table, Augmented model (Quintiles)			
Augmented model	Winner	Loser	Total	Augmented model	Winner	Loser	Total
Winner	1609	861	2470	Winner	4757	2983	7740
Loser	963	1309	2272	Loser	3045	4766	7811
Total	2572	2170	4742	Total	7802	7749	15,551
Differential excess return above the benchmark and the peer-group between winners and losers from the augmented model, post-estimation period $t + 36$ m							
Winners–Losers	+ 52.46 bp			Winners–Losers	+ 58.56 bp		
Panel B: Large Cap Growth							
Contingency table, Augmented model (Deciles)				Contingency table, Augmented model (Quintiles)			
Augmented model	Winner	Loser	Total	Augmented model	Winner	Loser	Total
Winner	1967	909	2876	Winner	6167	2839	9006
Loser	773	1496	2269	Loser	2844	4898	7742
Total	2740	2405	5145	Total	9011	7737	16,748
Differential excess return above the benchmark and the peer-group between winners and losers from the augmented model, post-estimation period $t + 36$ m							
Winners–Losers	+106.93			Winners–Losers	+104.03		

Notes: The upper part of the table reports 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model. Winners (losers) are funds in the top (bottom) deciles and quintiles alphas in period $t - 36$ m and top (bottom) deciles and quintiles alphas in period $t + 36$ m (rolled one month forward). Lower part of the table reports returns (in bp p.a.) in excess of the benchmark and the peer-group in period $t + 36$ m for all winner and loser funds selected by our augmented model in period t.

of Table 8 lays out the results based on our new alpha sorts in the form of 2 × 2 contingency table limited to the top and bottom deciles and quintiles only.[15] The table includes only the funds that are in the top/bottom decile or quintile in period t and remain there in period $t + 1$. It does not include funds that move from top/bottom to any other decile or quintile. As previously, Panel A corresponds to Large Cap Value funds and Panel B to Large Cap Growth funds. In the Appendix, we present full set of results for all deciles (quintiles) and 10 × 10 (5 × 5) contingency tables based on (i) our new augmented model alphas and (ii) Carhart alphas in Table A1 (Table A2). The results from the CTs in Table 8 confirm our previous findings that when selecting winner and loser funds using our new augmented model the persistence in performance exists (χ^2 significant at 1%) and that it is driven by winner funds. The slight exception is the persistence similarly driven by both winners and losers among quintiles of Large Cap Value funds in Panel A of Table 8. In the lower part of Panel A, we demonstrate that Large Cap Value funds selected in the top decile (quintile) in period t, generate over 50bps p.a. higher returns than bottom decile (quintile) funds once the returns of benchmark and the peer group are accounted for. Equivalent differential excess returns between the top and bottom decile (quintile) of Large Cap Growth funds exceeds 100 bp p.a. (Panel B). In the Table A1 and A2 in the appendix, persistence in performance, significant at 1%, is confirmed for both our model and the Carhart model and it is driven by both winners and losers. However, note the following: the quintile (decile) portfolios are not ideal for comparative analysis between the two models, as a fund with negative alpha in period t and $t + 1$ may be classified as a winner if it is placed in the top quintile (decile) of funds if there is very few funds with positive alphas at the time, which does not match a definition of a 'true' winner according to our new model. Therefore, comparing the returns in excess of benchmarks and the peer-group of our model vs. Carhart in the post estimation, i.e. post quintile/decile sorting period ($t + 36$) would be more meaningful. Panel A (Panel B) of Table A1 in the appendix shows that Large Cap Value (Large Cap Growth) funds selected by our model in the top decile can generate 26.9 bp p.a. (33.6 bp p.a.) higher returns than those selected by the Carhart model. Similarly, bottom decile Large Value (Large Growth) funds selected by our model have 25.2 bp (33.1 bp p.a.) lower excess returns than the funds selected by Carhart in the period $t + 36$ m, demonstrating that our model is more successful in framing future winners. Similar can be concluded from quintile fund sorts in Table A2 in appendix.

3.3.4. Cross-sectional regressions

In this section, we present results from two cross-sectional regressions to formally confirm that (1) alphas from our new model and (2) t-statistics of those alphas estimated in period t (i.e. $t-36$ m) explain fund's average excess returns above the benchmark and the peer-group in the subsequent period $t+1$ (i.e. in 36 months following the estimation of alphas):

$$Excess\ return_{i,t+1} = a_i + b_i Performance_{i,t} + u_{i,t} \tag{10}$$

where $Performance_{i,t}$ is either our new augmented alpha estimated in period t, $\alpha_{\mathrm{i,AGM,t}}$, or the t-statistics of that alpha, estimated using $t-36$ months of data; and $Excess\ return_{i,t+1}$ represents average excess return (in excess of self-reported benchmark and peer-group) of the fund over the 36 months following the estimation of augmented alpha. Model run with t-statistics of alpha serves as additional robustness test as some of the higher (lower) alphas may not necessarily be statistically significant,[16] but we expect that the increase in statistical significance of alphas will improve fund's future excess returns.

Panel A (Large Value) and Panel C (Large Growth) of Table 9, report the impact of our new augmented alpha in period t (estimated using $t-36$ m returns) on fund's excess returns in the subsequent 36 months ($t+36$). Panel B (Large Value) and Panel D (Large Growth) report results from the second model where our new augmented alpha in period t is replaced with the value of its t-statistics. We present results for all the funds together and split the funds across different benchmarks they report.

We find that for the whole sample and the sub-samples of funds benchmarking against typical passive indices for Large Value and Large Growth categories – such as Russell 1000 Value, Russell 3000 Value, Russell 1000 Growth and Russell 3000 Growth and S&P500 – the link between the past alphas (t-stats of alphas) and future excess returns is positive and significant at 1% level,[17] as per Table 9. This supports all our prior results, which establish that our new, augmented alpha model enables investors to successfully identify funds with superior future returns, in excess of benchmark and the peer-group.

3.3.5. Holding period 24 months

To test the robustness of our results to the length of the holding period, we re-create 2 × 2 contingency table corresponding to Table 4 in the paper, with focus on our new augmented model. Specifically, in Table 10, winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t-36$ m and positive (negative) alphas in period $t+24$ m (rolled forward by one month). The results are qualitatively the same as previously reported in Table 4. When using alphas from our new augmented model to sort the funds, our model identifies greater number of winners (6651 in Panel A and 6448 in Panel B) than losers. We find that W/W funds are the drivers of significant persistence in performance in Large Cap Value category at the 24 months horizon: there are over 1200 more W/W funds than L/L funds in Panel A (χ^2 significant at 1%). In Panel B, as in Table 4, the number of W/W and L/L funds is more even, but considerably larger than W/L and L/W groups – implying persistence in both W/W and L/L funds within the Large Cap Growth category (Panel B), χ^2 significant at 1%. This shows that our new, augmented model which accounts for both peer-group and the benchmark can identify persistence in performance even at shorter horizon of 24 months.

Similarly, Table 11 presents a 2 × 2 contingency table corresponding to our augmented model and method from Table 6 in the paper, where only funds with significant alphas in both the sorting and holding periods are considered for the construction of the table. Thus, in Table 11, winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t-36$ m and positive (negative) *and significant* alphas in period $t+24$ m (with one month rolling window). Our proposed augmented model continues to exhibit strong ability to select persistent winners in Large Cap Value Category (Panel A), and persistent winners and losers in Large Cap Growth Category (Panel B), even when 24-month holding period is used and only significant alphas are considered (χ^2 significant at 1% level). These results are qualitatively the same as those in Table 6 that correspond to 36-month holding period.

3.3.6. Small cap value and small cap growth categories

We have initially chosen Large Cap Value and Large Cap Growth categories as categories that contain larges number of funds and cover two distinct investment styles (value and growth). Our methodology can be applied

Table 9. The impact of past performance on subsequent excess returns.

Panel A: Large Cap Value, relationship between new alphas and future excess returns					
Benchmark Coefficient	ALL	Russel 1000 Value	Russel 3000 Value	SP500	Others
Performance in t,	0.0697***	0.0349***	0.0944***	0.0873***	−0.0066
α_{AGM}	(18.84)	(7.89)	(8.29)	(9.54)	(−0.55)
Constant	−0.0012***	−0.0009***	−0.0006***	−0.0019***	−0.0017***
	(−119.06)	(−85.53)	(−15.74)	(−69.05)	(−49.17)
Observations	87, 109	55, 014	6108	17, 893	8094
Panel B: Large Cap Value, relationship between t-stats of new alphas and future excess returns					
Benchmark Coefficient	All	Russel 1000 Value	Russel 3000 Value	SP500	Others
Performance in *t*,	0.0002***	0.0001***	0.0002***	0.0003***	0.0001
T-stat of α_{AGM}	(21.17)	(10.56)	(6.40)	(10.40)	(0.88)
Constant	−0.00115***	−0.0009	−0.00056***	−0.0019	−0.0017
	(−119.38)	(−85.97)	(−15.10)	(−68.44)	(−49.02)
Observations	87,109	55,014	6108	17,893	8094
Panel C: Large Cap Growth, relationship between new alphas and future excess returns					
Benchmark Coefficient	All	Russel 1000 Growth	Russel 3000 Growth	SP500	Others
Performance in *t*	0.06785***	0.02340***	0.0475***	0.1486***	0.03068
(α_{AGM})	(19.62)	(5.06)	(3.08)	(23.72)	(3.54)
Constant	−0.0010***	−0.0010***	−0.0004***	−0.0014***	−0.0005
	(−90.64)	(−68.39)	(−4.86)	(−69.88)	(−13 − 30)
Observations	90,464	41,888	2.754	31,984	13,838
Panel D: Large Cap Growth, relationship between t-stats of new alphas and future excess returns					
Benchmark Coefficient	All	Russel 1000 Growth	Russel 3000 Growth	SP500	Others
Performance in *t*	0.0003***	0.0017***	0.0006***	0.0004***	0.0002***
(T-stat of α_{AGM})	(29.26)	(14.55)	(10.53)	(23.92)	(6.84)
Constant	−0.0010	−0.0010***	−0.0006***	−0.0014***	−0.0005***
	(−90.37)	(−66.75)	(−7.88)	(−69.56)	(−13.92)
Observations	90,464	41,888	2.754	31,984	13,838

Notes: The table reports results from two cross sectional regressions for Large Cap Value funds (Panel A and B) and Large Cap Growth funds (Panel C and D). Panel A and Panel C, report the impact of our new augmented alpha in period *t* (estimated using $t-36$ m returns) on fund's excess returns in the subsequent 36 months ($t+36$). Panel B and D report results from the second model where our new augmented alpha in period *t* is replaced with the value of its t-statistics. The table reports the coefficients from each model and the number of observations, split by funds benchmarks in each panel. The values in parentheses used to gauge significance of coefficients are t-statistics. ***, **, and * denote significance at 1%, 5% and 10% level respectively.

to any category (peer-group) of funds. To test the robustness of our results to the choice of the peer-group, we apply our methodology to the Small Cap Value and Small Cap Growth categories. There are 187 unique Small Cap Value Funds and 369 unique Small Cap Growth funds in our sample following the same data cleaning process as for the Large Cap categories described in Section 2.1. Table 12 validates our earlier results from Table 4 and demonstrates that our new model is successful in identifying future winners, capable of outperforming their peer groups and self-reported benchmarks within the small cap styles. As in Table 4, W/W (L/L) category is defined as funds with positive (negative) and significant (10%) augmented alphas in the performance estimation period ($t-36$ m) and positive (negative), but not necessarily significant alphas in the subsequent 36 months ($t+36$).

In Panel A of Table 12, Small Cap Value W/W funds outnumber any other category in this 2 × 2 CT, once again highlighting our model's success in isolating persistent winners. The persistence of winners is even more pronounced within the Small Cap Growth peer-group, in Panel B, where the number of W/W funds that our model selects is nearly double the number of L/L funds, and larger than the number of funds in any other category. χ^2 significant at 1% level in both panels.

We now turn our focus towards funds that are identified as winners (losers) with positive (negative) and significant benchmark- and peer-group-adjusted alphas in the estimation period, that continue to generate positive (negative) *and statistically significant* alphas (at 10% level) in the subsequent 36-month period. The 2 × 2

Table 10. 2X2 Contingency tables augmented (new) model, 24 months holding period.

Contingency Table 2 × 2, Augmented model							
Panel A: Large-Cap Value				Panel B: Large-Cap Growth			
	Winner	Loser	Total		Winner	Loser	Total
Winner	4200	2451	6651	Winner	4009	2439	6448
Loser	2057	2929	4986	Loser	1912	4014	5926
Total	6257	5380	11,637	Total	5921	6453	12,374

Notes: Panel A (Large Cap Value funds) and Panel B (Large Cap Growth funds) of the table report 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t-36$ m and positive (negative) alphas in the holding period $t+24$ m (rolled one month forward).

Table 11. 2X2 Contingency tables augmented (new) model, 24 months holding period, significant alphas only.

Panel A: Large-Cap Value				Panel B: Large-Cap Growth			
	Winner	Loser	Total		Winner	Loser	Total
Winner	845	196	845	Winner	948	345	1293
Loser	246	617	246	Loser	195	977	1172
Total	1091	813	1091	Total	1143	1322	2465

Notes: Panel A (Large Cap Value funds) and Panel B (Large Cap Growth funds) of the table report 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t-36$ m and positive (negative) *and significant* alphas (10%) in the holding period $t+24$ m (rolled one month forward).

Table 12. Small Cap Value and Growth, 2X2 contingency tables augmented (new) model.

Contingency table 2 × 2, Augmented model							
Panel A: Small-Cap Value				Panel B: Small-Cap Growth			
	Winner	Loser	Total		Winner	Loser	Total
Winner	1048	956	2,004	Winner	3459	2159	5618
Loser	743	884	1627	Loser	1489	1862	3351
Total	1791	1840	3631	Total	4948	4021	8969

Notes: Panel A (Small Cap Value funds) and Panel B (Small Cap Growth funds) of the table report 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t-36$ m and positive (negative) alphas in the holding period $t+36$ m (rolled one month forward).

contingency tables, presented in Table 13, illustrate that our results remain qualitatively the same as in Table 6 when we exclude all funds with insignificant alphas in $t+36$ m: persistence in performance stems from W/W category within both, Small Cap Value (Panel A) and Small Cap Growth (Panel B) peer-groups.

Overall, the results from the robustness tests support our claim that our new, augmented model enables investors to successfully isolate funds that not only historically outperformed their self-declared benchmark and peer-group, but continue to do so in the future.

Table 13. Small Cap Value and Growth, 2X2 contingency tables augmented (new) model, significant alphas only.

Contingency table 2 × 2, Augmented model							
Panel A: Small-Cap Value				Panel B: Small-Cap Growth			
	Winner	Loser	Total		Winner	Loser	Total
Winner	185	178	363	Winner	690	256	946
Loser	84	128	212	Loser	377	553	930
Total	269	306	575	Total	1067	809	1876

Notes: Panel A (Small Cap Value funds) and Panel B (Small Cap Growth funds) of the table report 2 × 2 contingency tables to assess performance persistence based on alphas from our augmented model and on the standard Carhart model. Winners (losers) are funds with positive (negative) and significant (at 10%) alphas in period $t - 36$ m and positive (negative) *and significant* alphas (10%) in the holding period $t + 36$ m (rolled one month forward).

4. Conclusions

Standard Fama-French-Carhart models are widely used in the mutual fund performance literature and inform investors on funds' performance given the market, size, value and momentum risk factors. However, they do not reveal how funds fare against their benchmark or the peer group – a metric commonly referred to in funds' communications with investors. In this paper, we develop a new approach for assessing the performance of mutual funds that reflects the performance reporting style of fund performance in the industry. The model we propose alters the standard Carhart (1997) model to control for the benchmark performance *and* the performance of funds peer-group. Thus, it accounts for the benchmark by augmenting the left-hand side of the Carhart model, and for the commonalities of the peer group beyond those already embedded in the benchmark by adding 'peer-group factor' on the right-hand side of the Carhart equation. We illustrate the application of the proposed methodology using the group of Large Cap Value and Large Cap Growth active equity US mutual funds. Our results demonstrate that our model successfully identifies winner funds that outperform the benchmark and the peer-group three years ahead, while the standard Carhart model shows persistence among the loser funds only. The results are robust to different specifications of contingency tables, shorter holding period window (24 months), and alternative investment styles (namely, Small Cap Value and Small Cap Growth). To conclude, this paper sheds new light on Carhart's model underestimation of performance of winner funds and provides a solution for rectification that is of high interest to investors and practitioners. This is of particular importance in downward markets where all the funds in the peer group and the benchmark may be generating negative performance, but the fund that outperforms both would be the preferred one. The model can be applied to any group of equity funds and extended to other asset classes such as fixed income and hedge funds. In terms of further research, the relationship between performance using the model from this paper and funds' operational fund characteristics such as size or expense ratios can be examined, in line with Angelidis, Babalos, and Fessas (2021).

Notes

1. For detailed review of literature on persistence in mutual fund performance see Cuthbertson, Nitzsche, and O'Sullivan (2010) and Cremers, Fulkerson, and Riley (2019).
2. For evidence see Mateus, Mateus, and Todorovic (2016), Cremers, Petajisto, and Zitzewitz (2012, 2019), Elton and Gruber (2020).
3. Morningstar categorises funds into nine combinations of size (small, medium, large cap) and style (value, blend, growth) categories based on characteristics of their portfolio holdings.
4. There are nine (9) Morningstar categories in total: Large Cap Value, Large Cap Blend and Large Cap Growth; Mid Cap Value, Mid Cap Blend and Mid Cap Growth; and Small Cap Value, Small Cap Blend and Small Cap Growth.
5. We follow a number of papers referred to in our study that all assess mutual fund performance using monthly data: Hunter et al. (2014), Chinthalapati, Mateus, and Todorovic (2017), Mateus, Mateus, and Todorovic (2016, 2019a, 2019b, 2019c, 2023), Angelidis, Babalos, and Fessas (2021), Choi and Zhao (2021), Cuthbertson, Nitzsche, and O'Sullivan (2022, 2023).

6. For instance, in market downturns when both the index and funds within a group perform poorly, but a fund still performs less poorly than the index or the peer group. We will illustrate this with an example in Section 3.1 of the paper.
7. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
8. Note that $R_{bi,t}$ is the return of a fund's self-reported prospectus benchmark. For instance, in the Large Cap Value peer group, 91% of funds reports Russell 1000 Value, Russell 3000 Value or S&P500 as their benchmark, there is in total 24 benchmarks reported in this category. Similarly, there is a total of 34 benchmarks in Large Cap Growth peer-group category, with majority of funds reporting Russell 1000 Growth, Russell 3000 Growth and S&P500 as their benchmarks. To calculate benchmark-adjusted return for each fund *i*, we use their respective self-reported benchmark.
9. $R_{i,t} - R_{b,t} = \alpha_{i,b} + \beta_{i,b,M}(R_{M,t} - R_{f,t}) + \beta_{i,b,SMB}SMB_t + \beta_{i,b,HML}HML_t + \beta_{i,b,WML}WML_t + e_{i,b,t}$; where $\alpha_{i,b}$ is benchmark-adjusted alpha of fund *i*, $e_{i,b,t}$ is the difference between fund's and fund benchmark's idiosyncratic error terms with respect to the four factors and $\beta_{i,b,M}, \beta_{i,b,SMB}, \beta_{i,b,HML}, \beta_{i,b,WML}$ are the differences between the factor loadings of fund *i* and their primary prospectus benchmark index *b*. The factors are defined as per Equation (1).
10. Results available on request.
11. We require minimum of 30 months of data for each fund to be included in 36-monthly period 't' or 't+1' in all contingency tables we report.
12. We conduct robustness test where we define W/W funds as those that have $\alpha_t > 0$, and $\alpha_{t+1} > 0$, both statistically significant at 10% (based on a t-test from rolling regressions). The results are presented in Section 3.3.1 of the paper.
13. χ^2 significant at 1%.
14. All alphas are estimated as described in Section 3.2.
15. Note that the Winner/Loser funds in Table 8 are those that are in top decile (quintile) in period t and in the bottom decile (quintile) in period $t+1$ and vice versa for Loser/Winner funds in period $t/t+1$.
16. Approach used in Hunter et al. (2014) and Mateus et al. (2019b).
17. Less pronounced relationship is present in funds benchmarking against other, less suitable benchmarks for Large Cap Value and Large Cap Growth category. Benchmark mis-specification is discussed in Sensoy (2009) and Mateus, Mateus, and Todorovic (2019c).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Cesario Mateus is a Full Professor of Finance at Aalborg University. His research interests include corporate finance (capital structure, cash holdings, taxation, private equity, etc.), corporate governance (regulation, supervisory boards, and banks' risk-taking), asset management (mutual funds' performance and style investments) and SMEs and Family Businesses. He has worked at various universities in various countries, including Bayes Business School, Universitat Pompeu Fabra, University of Melbourne, Catholic University of Lisbon, Grenoble Ecole de Management, University of Southampton, and Aston Business School. He is the director of MSMEs and Family Businesses Research Lab, founder of FEB-RN, and Editor of Investment Analysts Journal.

Irina B. Mateus is an Associate Professor in Finance at Aalborg University. She is an active researcher, research interests include corporate finance (cash holdings, capital structure), asset management (mutual funds' performance and style investments), and capital investments (private equity and venture capital). Her research has been published in high-quality international peer-reviewed journals such as the European Journal of Finance, the Journal of Asset Management, the Quarterly Review of Economics and Finance, the International Review of Financial Analysis, among others. She is a member of the MSMEs and Family Businesses Research Lab and the Finance, Economics and Banking Research Network (FEB-RN).

Natasa Todorovic is a Associate Professor at Bayes Business School, where she obtained her PhD. Her research, which focuses on Asset Management, has been published in a number of peer-reviewed finance journals and presented at numerous leading conferences. Her most recent research interest lies in the area of improvement of mutual fund performance evaluation. She serves as an Associate Editor at The European Journal of Finance and a regular referee at a selection of peer reviewed journals. She has received several awards for her excellence in teaching. Natasa's links to finance industry and professional bodies include various posts in executive training as well as the Chief Examiner for Fund Management post at the Chartered Institute for Securities and Investment (London). Most recently, Natasa collaborated with The World Bank and Milken Institute to create a custom-made public financial asset management programme for future leaders of sovereign wealth funds, public pension funds and central banks in developing countries.

References

Angelidis, T., V. Babalos, and M. Fessas. 2021. "The Economic Gain of Being Small in the Mutual Fund Industry: U.S. and International Evidence." *International Review of Financial Analysis* 77:101852.

Angelidis, T., D. Giamouridis, and N. Tessaromatis. 2013. "Revisiting Mutual Fund Performance Evaluation." *Journal of Banking & Finance* 37 (5): 1759–1776.
Berk, J. B., and J. H. van Binsbergen. 2015. "Measuring Skill in the Mutual Fund Industry." *Journal of Financial Economics* 118 (1): 1–20.
Cai, B., T. Cheng, and Y. Cheng. 2018. "Time-varying Skills (Versus Luck) in U.S. Active Mutual Funds and Hedge Funds." *Journal of Empirical Finance* 49:81–106.
Carhart, M. M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52:57–82.
Chan, L. K. C., S. G. Dimmock, and J. Lakonishok. 2009. "Benchmarking Money Manager Performance: Issues and Evidence." *Review of Financial Studies* 22 (11): 4553–4599.
Chen, Z., and P. J. Knez. 1996. "Portfolio Performance Measurement: Theory and Applications." *Review of Financial Studies* 2:511–555.
Chinthalapati, V. L., C. Mateus, and N. Todorovic. 2017. "Alphas in Disguise: A new Approach to Uncovering Them." *International Journal of Finance and Economics* 22 (3): 234–243.
Choi, J. J., and K. Zhao. 2021. "Carhart (1997) Mutual Fund Performance Persistence Disappears Out of Sample." *Critical Finance Review* 10 (2): 263–270.
Clare, A., K. Cuthbertson, D. Nitzsche, and N. O'Sullivan. 2021. "How Skilful are US Fixed-Income Fund Managers?" *International Review of Financial Analysis* 74:101673. https://doi.org/10.1016/j.irfa.2021.101673.
Cremers, K. J. M., J. A. Fulkerson, and T. B. Riley. 2019. "Challenging the Conventional Wisdom on Active Management: A Review of the Past 20 Years of Academic Literature on Actively Managed Mutual Funds." *Financial Analysts Journal* 75 (4): 8–32.
Cremers, M., A. Petajisto, and E. Zitzewitz. 2012. "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation." *Critical Finance Review* 2:1–48.
Cuthbertson, K., D. Nitzsche, and N. O'Sullivan. 2010. "Mutual Fund Performance: Measurement and Evidence." *Financial Markets, Institutions and Instruments* 19 (2): 95–187.
Cuthbertson, K., D. Nitzsche, and N. O'Sullivan. 2022. "Mutual Fund Performance Persistence: Factor Models and Portfolio Size." *International Review of Financial Analysis* 81:102133.
Cuthbertson, K., D. Nitzsche, and N. O'Sullivan. 2023. "UK Mutual Funds: Performance Persistence and Portfolio Size." *Journal of Asset Management* 24:284–298.
Elton, E. J., and M. J. Gruber. 2020. "A Review of the Performance Measurement of Long-Term Mutual Funds." *Financial Analysts Journal* 76 (3): 22–37.
Fama, E. F., and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33:3–56.
Fletcher, J., and D. Forbes. 2002. "An Exploration of the Persistence of UK Unit Trust Performance." *Journal of Empirical Finance* 9 (5): 475–493.
Hunter, D., E. Kandel, S. Kandel, and R. Wermers. 2014. "Mutual Fund Performance Evaluation with Active Peer Benchmarks." *Journal of Financial Economics* 112 (1): 1–29.
Matallin-Saez, J. C. 2007. "Portfolio Performance: Factors or Benchmarks." *Applied Financial Economics* 17 (14): 1167–1178.
Mateus, I. B., C. Mateus, and N. Todorovic. 2016. "UK Equity Mutual Fund Alphas Make a Comeback." *International Review of Financial Analysis* 44: 98–110.
Mateus, I. B., C. Mateus, and N. Todorovic. 2019a. "Review of new Trends in the Literature on Factor Models and Mutual Fund Performance." *International Review of Financial Analysis* 63:344–354.
Mateus, I. B., C. Mateus, and N. Todorovic. 2019b. "Use of Active Peer Benchmarks in Assessing UK Mutual Fund Performance and Performance Persistence." *The European Journal of Finance* 25 (12): 1077–1098.
Mateus, I. B., C. Mateus, and N. Todorovic. 2019c. "Benchmark-adjusted Performance of US Equity Mutual Funds and the Issue of Prospectus Benchmarks." *Journal of Asset Management* 20 (1): 15–30.
Mateus, C., I. Mateus, and N. Todorovic. 2023. "Searching for Mutual Fund Winners? The Strategy is to Outbid Both, the Benchmark and the Peer Group." *Applied Economics* 56 (11): 1268–1282. https://doi.org/10.1080/00036846.2023.2175778.
Sensoy, B. A. 2009. "Performance Evaluation and Self-Designated Benchmark Indexes in the Mutual Fund Industry." *Journal of Financial Economics* 92 (1): 25–39.
Teo, M., and S. J. Woo. 2001. "Persistence in Style-Adjusted Mutual Fund Returns." Unpublished Paper, Harvard University. SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 291372.

Appendix

Table A1. Decile portfolios, 10 × 10 contingency table.

Panel A; Large Cap Value

Large Value, the new model

	Top	9	8	7	6	5	4	3	2	Bottom	Total
Top	1609	1153	933	722	722	785	636	810	1207	861	8929
9	1080	915	830	866	1062	925	796	881	1130	659	8689
8	800	1117	1071	1061	985	1125	792	763	858	740	8770
7	889	896	1153	791	932	930	980	701	875	730	8739
6	716	932	772	954	836	908	1025	806	890	919	8657
5	813	808	814	836	849	959	1040	993	738	813	8764
4	764	753	809	1121	867	878	970	863	789	847	8799
3	598	696	727	877	918	704	941	1185	737	787	8712
2	681	695	846	856	872	835	704	892	765	1120	8721
Bottom	963	706	768	597	668	785	820	858	698	1309	8599
Total	8913	8671	8723	8681	8711	8752	8704	8752	8687	8785	87,379

Large Value, Carhart model

	Top	9	8	7	6	5	4	3	2	Bottom	Total
Top	1913	1123	850	768	730	656	690	713	774	710	8936
9	1319	958	824	838	705	796	942	732	886	683	8683
8	685	908	1040	1077	813	821	747	981	845	851	8768
7	757	810	1134	972	1164	910	910	803	691	588	8739
6	791	901	1146	959	812	871	925	857	696	706	8664
5	613	854	847	944	952	1021	903	992	824	814	8764
4	628	969	818	811	792	839	1161	923	1046	814	8801
3	605	759	700	918	1039	943	956	976	930	872	8698
2	715	679	709	767	953	970	853	927	995	1160	8728
Bottom	904	737	673	718	701	936	682	841	1025	1381	8598
Total	8930	8698	8741	8772	8661	8763	8769	8745	8712	8588	87,379

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model; Post Estimation Period: $t + 36$ m

Top	9	8	7	6	5	4	3	2	Bottom
+26.9 bp	+21.2 bp	+28.6 bp	+6.8 bp	−30.4 bp	+2.5 bp	+6.67 bp	−16.6 bp	−21.7 bp	−25.2 bp

Panel B: Large Cap Growth											
Large Growth, the new model											
	Top	9	8	7	6	5	4	3	2	Bottom	Total
Top	1967	1361	953	861	716	606	607	581	617	909	8900
9	1487	1352	1024	901	845	786	700	630	686	627	9074
8	1076	1106	1055	981	929	855	837	817	742	670	9032
7	857	947	929	960	973	912	950	937	889	670	9060
6	671	794	962	1063	1056	969	911	951	913	706	9094
5	547	697	933	1019	1025	1086	1027	990	955	815	8996
4	587	695	902	922	915	1007	1101	1051	1018	862	9024
3	520	708	790	867	930	1095	1094	1031	1026	971	9068
2	693	709	774	761	837	984	953	1080	1099	1174	9038
Bottom	773	669	746	689	770	794	870	964	1129	1496	9178
Total	9178	9038	9068	9024	8996	9094	9060	9032	9074	8900	90,464

Large Growth, Carhart											
	Top	9	8	7	6	5	4	3	2	Bottom	Total
Top	1600	1081	931	838	714	664	669	722	767	1, 192	9178
9	1022	1100	985	918	913	913	860	832	742	753	9038
8	950	1030	1,031	971	926	965	917	874	803	601	9068
7	825	1003	950	928	982	951	962	890	837	696	9024
6	806	903	950	951	978	985	986	873	838	727	8997
5	732	861	943	1023	1031	954	1043	915	849	742	9093
4	739	801	860	935	990	907	974	1013	948	893	9060
3	661	712	804	860	928	1053	944	976	1077	1017	9032
2	796	744	763	848	800	959	933	1100	1109	1022	9074
Bottom	1016	772	820	722	703	713	741	806	1073	1534	8900
Total	9147	9007	9037	8994	8965	9064	9029	9001	9043	9177	90,464

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model; Post Estimation Period: $t + 36$ m								
Top	9	8	7	6	5	4	3	2	Bottom
+33.6 bp	+38.2 bp	+5.5 bp	+12.1 bp	−8.9 bp	−20.4 bp	−19.8 bp	−3.4 bp	−18.3 bp	−33.1 bp

Notes: The Table presents 10 × 10 contingency table. Winners/Winners (Losers/Losers) are funds in the top (bottom) deciles based on alphas in period t (estimated using $t - 36$ months), and continue to be in the top (bottom) deciles are based on alphas in period $t + 36$ m (rolled one month forward, over 336 rolling periods). Lower part of the table reports returns (in bp p.a.) difference in returns (in bp p.a.) among the augmented and the Carhart models in excess of the benchmark and the peer-group in period $t + 36$ m for all winner and loser funds.

Table A2. Quintile portfolios, 5 × 5 contingency table.

Panel A: Large Cap Value						
Large Value, the new model						
	Top	4	3	2	1	Total
Top	4757	3351	3404	3123	2983	17, 618
4	3702	4076	3499	3236	2996	17, 509
3	3269	3376	3552	3864	3360	17, 421
2	2811	3534	3840	3959	3367	17, 511
Bottom	3045	3067	3168	3274	4766	17, 320
Total	17, 584	17, 404	17, 463	17, 456	17, 472	87, 379

Large Cap Value, Carhart model						
	Top	4	3	2	1	Total
Top	5313	3280	2887	3077	3062	17, 619
4	3160	4223	3708	3441	2975	17, 507
3	3159	3896	3656	3677	3040	17, 428
2	2961	3247	3613	4016	3662	17, 499
Bottom	3035	2867	3560	3303	4561	17, 326
Total	17, 628	17, 513	17, 424	17, 514	17, 300	87, 379

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model; Post Estimation Period: $t + 36$ m				
Top	4	3	2	Bottom
+24 bp	+17.7 bp	−13.9 bp	−4.9 bp	−23.4 bp

Panel B: Large Cap Growth						
Large Growth, new model						
	Top	4	3	2	1	Total
Top	5167	3739	2953	2518	2839	18, 216
4	3986	3925	3669	3541	2971	18, 092
3	2709	3977	4136	3879	3389	18, 090
2	2510	3481	3947	4277	3877	18, 092
Bottom	2844	2970	3385	3877	4898	17, 974
Total	18, 216	18, 092	18, 090	18, 092	17, 974	90, 464

Large Growth, Carhart						
	Top	4	3	2	1	Total
Top	4803	3672	3204	3083	3454	18, 216
4	3808	3880	3824	3643	2937	18, 092
3	3302	3867	3948	3817	3156	18, 090
2	2913	3459	3878	3907	3935	18, 092
Bottom	3328	3153	3175	3580	4738	17, 974
Total	18, 154	18, 031	19, 029	18, 030	18, 220	

Difference in post-estimation excess return (beyond the benchmark and the peer-group) between winner (loser) funds selected by the new augmented model vs. Carhart model; Post Estimation Period: $t + 36$ m				
Top	4	3	2	Bottom
+35.9 bp	13.9 bp	−12,7 bp	−11.6 bp	−25.6 bp

Notes: The table presents 5×5 contingency table. Winners/Winners (Losers/Losers) are funds in the top (bottom) quintiles based on alphas in period t (estimated using $t - 36$ months), and continue to be in the top (bottom) deciles are based on alphas in period $t + 36$ m (rolled one month forward, over 336 rolling periods). Lower part of the table reports returns (in bp p.a.) difference in returns (in bp p.a.) among the augmented and the Carhart models in excess of the benchmark and the peer-group in period $t + 36$ m for all winner and loser funds.