Does equity mutual fund factor-risk-shifting pay off? Evidence from the US

Cesario Mateus*
University of Aalborg Business School, Aalborg, Denmark
cmateus@business.aau.dk

Sohan Sarwar
University of Greenwich Business School, London, UK
g.sarwar@greenwich.ac.uk

Natasa Todorovic
Bayes Business School, City, University of London, UK
natasha.todorovic.1@city.ac.uk

Abstract

In this paper, we assess the relationship between risk-shifting of mutual funds, measured as benchmark-adjusted factor-based investment style change following a structural break, and their risk-adjusted performance. We isolate only the breaks in style risk beyond those embedded in the funds' benchmark index to eliminate any natural style risk changes resulting from varying company fundamentals over time. We group style risk changes into extreme (style rotation), moderate (style drifting), and weak (style-strengthening/weakening) and assess which investment style category is most profitable to shift in to and out of. Our findings show that funds that exhibit breaks generate overall better risk-adjusted performance than those that do not. Funds that are most successful in risk-shifting have both statistically and economically distinct risk-adjusted performance, make shifts towards small/large/value/growth style combinations rather than mid-cap and blend style, exhibit breaks less frequently and has more moderate risk-shifts than funds that are unsuccessful.

The European Journal of Finance, forthcoming

*corresponding author

Keywords: Structural breaks; style risk; risk shifting; performance; mutual funds

JEL: G11, G12, G2

1. Introduction

During the 1970s and 1980s there has been a mount of evidence in support of the investment style consistency argument, where value style stocks are found to outperform growth and small-cap stocks are found to outperform large caps over decades[1]. A simple implication of these findings would suggest that staying true to the value and/or small cap style should pay off in the long run. However, more recently, the evidence points at the disappearance of small size effect[2] and shows that in the recent decade it is actually growth stocks, not value, that exhibit superior performance[3]. In line with this evidence, the literature also shows that mutual fund portfolio managers who alter their portfolio holdings more frequently in pursuit of winner stocks drift away from their designated investment style (see e.g. Wermers, 2012). Note that the style drift does not necessarily have to be intentional – for instance, a portfolio of mid-cap stocks may eventually drift into a large cap-style if the values of assets increase over time. However, regardless of whether the change in portfolio characteristics is intended or not, its main implication is that it changes the risk of the portfolio (see Wermers, 2012, Huang et al., 2011, Brown and Harlow, 2004 for instance). Such alteration of risk, we believe, has most bearings on investors, particularly if it changes beyond that observed in a fund's benchmark index. Hence, in this paper, we examine to what extent US long-only mutual funds exhibit significant structural changes in their style risk exposures, beyond the changes in risk embedded in their self-reported prospectus benchmark. We also assess whether fund managers with more extreme risk shifts, resulting in change of fund's style classification, are able to yield stronger risk-adjusted performance than those with less extreme drift or those remaining true to their style.

Style drifts have been long documented in the literature on mutual fund performance. Brown & Goetzmann (1997), diBartolomeo & Wikowski (1997), Kim, Shukla and Tomas (2000), among others, report a significant mismatch between mutual funds' investment objectives and their style of investing. The link between the style drift and performance of mutual funds is not clear-cut. While some findings point that being style consistent leads to better absolute and relative performance (Brown, Harlow and Zhang, 2009), others document that allowing for style drifts implies better performance (e.g. Andreu, Sarto and Serrano, 2019, more recently),

[1] Examples can be found in Basu (1975), (1977) and (1983), Reinganum (1981), Banz (1981), Levis (1989), Reinganum (1992), Fama and French (1992), among other numerous studies
[2] Van Dijk (2013), Hou and Van Dijk (2019)
[3] https://www.morningstar.com/articles/1017342/value-vs-growth-widest-performance-gap-on-record

especially when it comes to small-cap funds, growth funds and funds with high alphas, even when extra trading costs are taken into account (Wermers, 2012).

According to Wermers (2012), style drift can be viewed as a change in portfolio characteristics over time (as first defined by Daniel and Titman, 1997) or as a shift in factor loadings on style risk factors, such as those in the Fama and French (1993) model. In the former approach (more recently used in Andreu et al, 2019), also known as holdings- or characteristics-based approach, to determine the degree of the drift one would have to know the exact holdings of funds and ideally the timings of the trades. According to SEC regulations in the US, mutual funds are required to report holdings up to 60 days *following* the quarter-end. What is not known is when the trading of the reported holding actually occurs or why the fund managers have altered the composition of their portfolio. This leads to a discrepancy between the fund returns inferred by the holdings and the returns actually earned. The phenomenon of adjusting the holdings prior to disclosure is known in mutual fund literature as 'window dressing', documented by Musto (1997, 1999) and Aggarwal, Gay and Ling (2014), among others. Such 'window dressing' not only misleads investors about fund performance but also causes alterations to fund risks (both upwards and downwards). Hence, inferring fund style from holdings can lead to inaccurate style and risk classification of a fund[4]. On the other hand, Davies, Fama and French (2001) argue that the link between portfolio characteristics and returns is limited to Daniel and Titman (1997) study period. The second strand of literature applies returns-based approach to style determination. To that end, for instance Davies (2001) and Hermann et al. (2016) define styles as exposure to Fama-French factors, and Annaert and Van Camplenhout (2007) apply return-based Sharpe (1992) style analysis approach. Finally, Brown, Harlow and Zhang (2009), use returns-based approach to supplement the holdings-based approach when assessing style consistency in mutual funds. Hence, due to the drawbacks of the holdings data, in this paper, we use returns-based approach and view fund risk shift as a change in size and style factor loadings from the standard Fama-French model, similar to Hermann et al. (2016).

We contribute to the literature on mutual fund style drifts and performance measurement. In spirit, and methodologically, our paper is closest to Annaert and Van Campenhout (2007) and Herrmann et al. (2016). The studies examine structural breaks in mutual fund's exposures to style risks using returns-based style analysis. However, while Anaert and Van Campenhout (2007) consider structural breaks that occur in Sharpe's style factor loadings for European

[4] More detailed discussion on the issue of 'window dressing' can be found in Section 2 of this paper.

funds that could be subject to benchmark specification error, we focus on the US equity fund risk-shifts beyond the passive shifts embedded in the fund's own prospectus benchmark, as indicated by Fama-french factor loadings. On the other hand, Hermmann et al. (2016) consider fund's tracking error, R-squared and changes in Fama-French factors from one quarter to the next to determine risk shifts. They consider absolute change in factor loadings as indicator of active shifts but do not take into account the sign or significance of the shift that we account for in our structural break analysis; and define passive shifts as keeping style exposure constant. Passive shifts refer to shifts that occur as a natural process of assets within the fund increasing/decreasing in e.g. size or P/E, not as a result of true fund-managers change in investment style. Wermers (2012) highlights that passive shifts significantly contribute to the overall drift in portfolio management style. To tackle this issue in our study, we consider benchmark-adjusted changes in style risk to be 'true' active risk shifts in our funds, given their prospectus benchmark. We define significant benchmark-adjusted style risk changes as breakpoints in benchmark-adjusted size and style factor betas, beyond the changes embedded in the funds' prospectus benchmark. To do this, we utilize the Angelidis, Giamouridis and Thessaromatis (2013) approach (AGT hereafter) to obtain funds' exposures to style risks in excess of (or below) the risk exposures of their benchmark. The AGT model adjusts the left-hand side of the standard Fama-French-Carhart equation by replacing the risk-free rate with the return of the benchmark, resulting in benchmark-adjusted loadings of funds (i.e. differences between a fund's and benchmark's style risk exposure). Hence, we capture breakpoints in benchmark-adjusted style risks, i.e. the changes in fund's style risk beyond the change embedded in the benchmark. To the best of our knowledge, this is the first paper that investigates the link between different degrees of *factor-based* benchmark-adjusted risk-shifting of mutual funds and the risk-adjusted-performance.

Our sample of funds comprises 1281 US long-only active equity mutual funds reporting S&P500 index as their primary prospectus benchmark, from January 1992 to December 2016. To examine benchmark-adjusted changes in style exposure, we apply Bai and Perron (2003) method to the AGT, benchmark-adjusted four-factor model. This allows us to identify (multiple) structural breaks in fund's factor loadings *in excess* of those of the given benchmark over time, at dates unknown prior to estimation. We do this to identify the funds which are truly active in their risk shifting (both upwards and downwards) and their change is not the result of a natural shift in style reflected in the benchmark index. To examine how each fund changes risk when a breakpoint does occur, we determine fund's investment style in the period before and after each benchmark-adjusted structural break utilizing the standard Fama and French (1993) model[5]. We place a fund into one of nine Morningstar style-box categories before and after each structural break in AGT-betas is identified. We identify three levels of intensity of risk shifting based on the type of change in the SMB and HML beta before vs. after the break: style rotation, style drift and style strengthening/weakening. Note that we define more extreme shifts those that lead to change is the sign and significance of beta coefficient, (i.e. the change in fund's style classification) not as those resulting in the largest absolute change in betas as in Herrmann et al., 2016. We discuss this further in Section 2.2.3.

We find that 367[6] funds in our sample experience at least one benchmark-adjusted break in style risk. Note that investors selecting a particular investment style have set risk constraints and return objectives matching that style, and they would expect a fund to be style consistent and true to their objectives at all times. In the context of AGT model, investors would expect having constant style exposure relative to the benchmark if the investment objectives are to be consistently met. Given that nearly 30% of funds in our sample deviate from this notion and exhibit breaks that could be causing a drift or change in investment style which is not embedded in investor's expectations – the effect of benchmark-adjusted risk shifting on subsequent fund performance warrants further investigation. There could be multiple reasons behind these shifts, from change in fund managers and their strategy, macroeconomic conditions, etc., and they are not subject of this research.

Our results show that funds that break exhibit better risk-adjusted performance than those with no true change in factor risk exposure during our sample periods. Further, funds that start as mid-cap or blend funds benefit most from shifting away from those categories. We show that on average, risk shifting beyond the benchmark leads to improvement in excess returns, reduction of standard deviation and improvement in Sharpe ratios in the period following a break-in risk. However, when it comes to standard Fama-French alphas or benchmark-adjusted (AGT) alphas, the improvement occurs following only around half of the total number of 'true' active breaks. We note that most of the funds shift towards 'safer' categories (mid-cap and blend). However, one of our findings with most implications for funds managers is that the greatest increase in alphas occur when the funds shift their risks towards more extreme size (small and large) and style (value and growth) categories. This leads us to believe that not all

[5] Note that the AGT model cannot provide information on fund's style, only information on whether a fund's factor risk exposure is higher/lower than that of the benchmark. Please refer to methodology for detail.

[6] Note that if we did not account for the benchmark and used standard Fama-French three-factor model to identify the breakpoints, there would be 972 funds with at least one structural break in the sample.

funds that undergo benchmark-adjusted shift in risk do so successfully. To this end, we find that performance of top and bottom decile of funds that have breaks in risks is statistically and economically very different. In addition, top decile funds resort more to moderate change in risk (drift) and shift more towards extreme style categories (small/large/value/growth) compared to the funds in the bottom decile. Top decile funds also exhibit less frequent structural breaks than bottom decile funds. These findings bare implications for investors looking to invest in a particular investment style and call for caution, as not all funds within the selected style will remain style-consistent and not every drift or shift to another style will be successful.

The remainder of this paper is organised as follows. Section 2 describes the data and methodology. Sections 3 and 4 provide analysis of the results. Section 5 summarizes the main findings and discusses the main implications of this study for investors and industry professionals.

2. Data and Methodology

2.1. Data

Our data set comprises 1281 US long-only active equity mutual funds that report the S&P 500 index as their primary prospectus benchmark in the Morningstar database. The sample period spans from January 1992 to December 2016. Fund returns inclusive of dividends and the information on primary prospectus benchmarks is from Morningstar. If all the funds in our sample followed the same investment style as the S&P500, they would all be placed in the Large Cap (in terms of size) - Blend (in terms of value/growth style) category in the Morningstar database – nonetheless, our results show they are not. This is in line with a number of papers (see Sensoy, 2009, for earlier and Mateus, Mateus and Todorovic, 2019b for most recent evidence) which document that funds do not necessarily follow the style of their primary benchmarks.

To establish a style category of a fund in month *t* and place it in the relevant Style Box, Morningstar uses the information on trailing fund holdings over the past 36 months, as reported by the fund itself. The Morningstar approach reflects the characteristics-based (holdings-based) approach for determining investment styles, illustrated in Daniel, Grinblatt, Titman and Wermers (1997). The approach requires the information on fund holdings, and it is based on the argument (see Daniel and Titman, 1997) that portfolio characteristics rather than factor loadings from the Fama-French model are better determinants of expected returns. In this paper, however, we infer a manager's investment style by utilising factor loadings from the Fama and French (1993) three-factor model, as seen in Davies (2001) and Herrmann et al. (2016). The reason for that is two-fold. First, Davies, Fama and French (2000) show that the advantages of the characteristics-based model are confined to Daniel and Titman (1997) study period. Second, the holdings data required for the application of characteristics-based approach has a significant flaw, as documented in recent literature. To that end, Aggarwal, Gay and Ling (2014) show that the regulation about disclosure of fund holdings leads to "window dressing" of information, causing discrepancies between a fund's actual returns and those obtained by replicating the disclosed portfolio holdings. In the US, SEC regulation requires funds to disclose their holdings quarterly within 60 days following quarter-end. "Window dressing" in the context of fund holdings refers to the behaviour where portfolio managers alter fund holdings prior to disclosure in an attempt to mislead investors by showing larger (smaller) holdings in winner (loser) assets. Additional evidence in this area suggests that portfolio managers do not necessarily alter only the perception of performance through "window dressing", but they tend to change portfolio risk characteristics by selling (buying) higher-risk (lower-risk) securities before the disclosure (Musto, 1997 and 1999; Morey and O'Neal, 2006). In addition, poor-performing funds, in particular, are found to invest in stocks that do not comply with fund's investment style and objectives, selling them before reporting holdings (Meier and Schaumburg, 2004)[7].

Hence, to assign an investment style to a fund in our sample, we apply the standard Fama-French (1993) three-factor model which enables us to place a fund into one of the nine size/style categories, as per Morningstar Style Box[8]. The debate about the accuracy of the Fama-French factors and omitted variables in the model has been the subject of finance literature for some time (see, for instance, Mateus, Mateus and Todorovic, 2019a for the review of recent literature on factor models, and Harvey, Liu and Zhu, 2016, for the discussion on the number of additional factors in literature). Despite the criticisms, there is no consensus as to which alternative model is better, so the three-factor model (together with Carhart, 1997) remains the most widely used benchmark model in the style investing literature.

[7] Additional evidence on "window dressing" can be found in Sias and Starks (1997), He, Ng, and Wang (2004), Ng and Wang (2004), Aggarval, Daniel and Naik (2011) among others.

[8] We explain the application of this method for fund-style categorization before and after the benchmark-adjusted breaks in more detail in the following section of the paper.

Further, when a passive benchmark generates outperformance in the standard Fama-French-Carhart factor model, a fund that simply replicates such a benchmark will appear to be outperforming[9]. Similarly, for these funds, any significant changes in style factor loadings of the benchmark will falsely indicate active risk-shifting in the fund. Therefore, to isolate only the true active changes in funds' factor loadings beyond those of the benchmark index, we assess structural changes in *benchmark-adjusted* size (SMB) and style (HML) betas for each fund. The methodology for determining benchmark-adjusted factor loadings and determining the presence of structural breaks is outlined in the following section.

2.2. Methodology

Our methodology is comprised of four steps: **1.** We start by determining structural breakpoints in the fund's benchmark-adjusted size and style betas, **2.** Next, we determine fund's size and style before vs. after the break, **3.** Then, we classify the levels of intensity of risk shift for each fund following each of its breaks, i.e. we determine whether the change is Fama-French size and style coefficient results in style rotation, style drift, or style strengthening/weakening; and **4.** Last, we test whether active risk-shifting beyond the benchmark pays off more than being more consistent and whether there are commonalities among funds with more successful benchmark-adjusted shifts. The four steps are detailed below.

2.2.1. Breakpoints in funds' benchmark-adjusted size and style betas

Similar to Annaert and Van Camenhout (2007), this study uses Bai-Perron's (2003) methodology to identify the endogenous structural breaks (either partial or pure) of mutual fund returns for any changes in fund's style factor loadings. Our method differs from Annaert and Van Campenhout (2007) in that we do not use daily mutual fund returns. We follow standard mutual fund performance literature - to avoid the noise in daily returns and short-lived breaks in this context– and utilise monthly returns data. Further, while their study examines time variation in Sharpe style factor loadings for European funds, where a choice of style index can drive fund's style determination; we focus on fund's Fama-French factor loadings amended by the loadings of a fund's self-reported benchmark in an attempt to isolate 'true' shifts in fund styles risks, beyond those of the fund's benchmark. As mentioned before, some changes in

[9] See Cremers et al. (2012) and Chinthalapati et al. (2017) for some evidence on non-zero benchmark alphas.

fund's factor loadings may occur naturally when, for instance, a small-cap stock becomes mid-cap. Since many of such stocks are part of the benchmark index too, those natural changes will be affecting factor loadings of the fund's benchmark as well. The importance of such 'passive' drifts within the overall style drift of the portfolio was documented by Wermers (2012). In the factor-based framework of this study, by applying a benchmark-adjusted model to identify structural changes in benchmark-adjusted size/style betas we dampen the effect of passive shifts and isolate those changes that occur as the result of 'true' active risk-shifting by fund managers.

As highlighted in the introduction of this paper, to determine the significant changes in fund's styles, beyond those of the benchmark index, we deploy Angelidis, Giamouridis and Thessaromatis (2013) approach (AGT) for estimating funds' benchmark-adjusted factor betas and benchmark-adjusted alphas. The approach can be viewed as the amendment of the standard three-factor model:

$$R_{i,t} - R_{Benchmark,t} = \alpha_i^* + \beta_{i1}^*\left(R_{M,t} - R_{f,t}\right) + \beta_{i2}^* SMB_t + \beta_{i3}^* HML_t + \varepsilon_{it}^* \qquad (1)$$

where $\mathrm{R_{i,t} - R_{Benchmark,t}}$ represents the amendment of the standard three-factor model (given in section 2.2.2. below) whereby the risk-free rate is replaced with the returns of the benchmark index (in our paper, a fund's self-reported benchmark index – the S&P500) to obtain the benchmark adjusted return of a mutual fund i in period t. α_i^* then represents fund's benchmark-adjusted alpha. $\left(R_{M,t} - R_{f,t}\right)$, is the market risk premium[10], SMB is the size factor obtained as a difference in returns between small-cap and large (big) cap firms and HML is the style factor obtained as the difference in returns between the firms with the high book-to-market (value firms) and low book-to-market ratio (growth firms), all as in Fama and French (1993). Hence, $\beta_{i1}^*, \beta_{i2,}^* \beta_{i3,}^*$ represent the difference between the fund's and benchmark's market, size and style factor loadings, i.e. benchmark-adjusted style risk exposure of a fund. In particular, $\beta_{i2,}^*\ \beta_{i3}^*$ show a fund's over/under-weighting of small/large or value/growth stocks relative to the self-reported benchmark index.

We then estimate the structural breaks in fund's benchmark-adjusted factor loadings. The literature on structural breaks is pioneered by Quandt (1958) and Chow (1960) who developed tests for a structural break for a single known date. To this end, we utilise Bai and Perron's

[10] US market risk premium represents the return (value-weighted) of all US firms in CRSP database listed on the NYSE, AMEX, or NASDAQ

(1998, 2003) methodology, as it considers partial structural change as well as pure structural change. In addition to Annaert and Van Campenhout (2007), the Bai-Perron method has been applied in the finance literature in Rapach and Wohar (2006), Paye, and Timmermann, (2006), Rapach, Strauss, and Zhou (2010), Lu and Perron (2010), for instance. To the best of our knowledge, this is the first time the method is utilized in the context of determining benchmark-adjusted risk shifting in mutual funds.

Table 1 outlines the breakdown of funds per number of structural breaks. It shows that 367 funds in our sample have at least one structural change in risk and the total number of regime changes in our sample is 1,558. The vast majority (286) of those funds have five structural breaks (which corresponds to 1,430 shifts in risk we analyse).

Table 1: Number of funds per structural break in AGT factor loadings

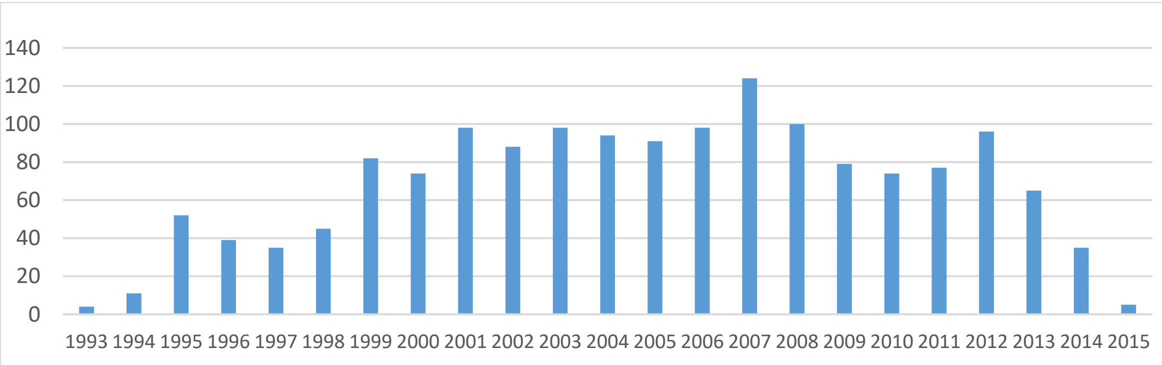
Table shows the number of funds per number of breaks and the number of breaks in total.

Number of breaks	Number of funds	Total breaks
1	55	55
2	12	24
3	7	21
4	7	28
5	286	1,430
Total	**367**	**1,558**

These results are in line with the results of many previous studies such as Annaert and Van Campenhout (2009), Brown et. al (2011) and Wermers (2012), among others, who document time-varying, inconsistent style exposures by mutual funds. However, in this study, by identifying structural changes in the benchmark-adjusted AGT model, we have eliminated a considerable number of what we consider false changes in size/style risk exposure, driven by changes in stock characteristics reflected in the market index used as a benchmark (S&P500). Specifically, should we have used the standard Fama-French three-factor model to identify the breakpoints instead of the AGT, there would have been 972 funds with at least one structural break(s) in size/style betas, a result much closer to the Annaert and Van Campenhout (2007) study that uses daily data and finds at least one break in all the funds in their sample. By accounting for the S&P500 index, we capture the element of risk shifting common for the stocks in the S&P 500 and are left with only true active risk changes beyond those implied by the index a fund uses as a benchmark. This results in 367 mutual funds with structural breaks and 1,558 individual breaks in total. We outline in Section 2.2.3 how we classify those benchmark-adjusted risk changes according to their intensity. This implies that the vast majority of funds in our sample (914) do not change the size/style risks beyond the changes in the benchmark index and remain what we consider in this study style consistent.

The 1,558 breakpoints in our sample are distributed across the years in our sample, but unsurprisingly we note particular concentration during the known crisis periods, with particular spike in breaks in anticipation of the start of the most recent financial crisis (2007/08), indicating that funds may change their style risk exposures to salvage their performance in the crisis period. We investigate this point further in Section 3.1.1. of the paper. Figure 1 presents this distribution of breaks during our sample period.

Figure 1: Number of breaks per year



2.2.2. The investment styles of funds

Following Davies (2001), we determine a fund's investment style according to the values and significance of the size and style coefficients obtained from the Fama-French (1993) three-factor model framework (FF3 hereafter). The FF3 model is given by:

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{i,M}\left(R_{M,t} - R_{F,t}\right)_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + e_t \qquad (2)$$

Where R_{it}, is the return of fund *i*, net of fees, R_F is the US 1 month Treasury bill; $R_{M,t} - R_{F,t}$ *SMB* and *HML* are Fama and French (1993) market risk premium, size and value (high minus low book-to-market company returns) factors respectively, as defined in equation (1). The data on US one-month Treasury bill and all Fama-French factors are from Kenneth French's website[11].

Note that while we apply AGT model to identify benchmark-adjusted change in risk, the model does not give us information on fund style, it only reveals whether the fund has greater or lower factor risk exposure relative to the benchmark, as per interpretation of equation (1). Therefore, to determine the investment style for each fund that exhibits a structural change in risk, the standard FF3 model is applied before (period t) and after (period t+1) each break occurs. Note that we do not impose minimum break length in Bai-Perron methodology in order not to influence results and miss important breaks. This implies that some of our break periods during which we estimate style of the fund may be short, implying a potential beta estimation error. We deal with that empirically throughout of the paper by applying robustness analysis with restricted sample not including breaks (funds with breaks) shorter than 24 months. Finally, for funds with no break in benchmark adjusted size/style exposure, the style is estimated for the overall sample period using equation (2).

Using the sign and significance of the SMB and HML factor loadings, in each risk regime, we classify the funds in our sample into one of the nine style categories from the Morningstar Style Box, obtained as combinations of three style categories (value, blend, and growth) and three size categories (small-, medium- and large-cap). The purpose of identifying fund's style exposure before and after the structural break in their benchmark-adjusted beta occurs is to identify the funds that have significantly shifted their risk before and after each break has occurred. Hence, if after identifying a breakpoint, the SMB coefficient is showing that the mutual fund's SMB beta has progressed from significantly positive before the break to significantly negative after it, for example, it implies this has fund changed its size exposure from small-cap to large-cap stocks significantly altering its risk profile. Similar can be said for the changes in coefficients associated with HML factor. The shifts in SMB and HML beta coefficients may result in a complete rotation in style as illustrated in this example, or some less extreme style drifts; all to be defined in the next section of the paper.

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

2.2.3. The intensity of risk-shifting (style change)

In this paper we define the level of intensity of risk shift not by looking at the absolute change in factor loadings (as in Herrmann et al., 2016) before and after the break but we account for the sign and the significance of the loadings. Consider, for instance, a fund that increases the beta by 0.3; if the beta remains of the same sign and significance, the fund is placed in the same Morningstar style (risk) category, so investor would not view such factor risk change as extreme. One the other hand, a change in betas of 0.3 that is accompanied by the change in sign and significance (say SMB beta changes from +0.1 to -0.8, both significant), implies that a manager changes exposure from small to large cap stocks, which for a fund branded as a 'small cap fund' is a cause for concern among investors.

Hence, once the structural break(s) for each fund has been established, we determine funds investment style by estimating the FF3 model in the regime prior to (period t) and post (period t+1) structural change. The level of change in the sign and significance of the estimated SMB and HML factor loadings between the regimes infers a different degree of shifting of the fund's size and style risk. To which extent the fund changes its risk exposure following the benchmark-adjusted break depends on the type of change in SMB and HML beta coefficients compared to the regime before the break. To this end, we group the funds according to the intensity of their benchmark-adjusted factor-risk-shifting into three categories, namely:

1) ***Style rotation funds***: those whose SMB or HML beta(s) remain significant before (time t) and after the break (time t+1), but change the sign, indicating that the fund has changed the style from one regime to another. Note that funds that exhibit more than one structural break can change style more than once over the sample period. The success of such a strategy where fund managers rotate between the styles has been well documented in the financial literature. Evidence on successful style rotation strategies in the US market can be found in Kao and Shumaker (1999) and Asness, Friedman and Liew (2000) among others and is not contained to the US market only[12]. However, the evidence also highlights that a typical mutual fund would have risk constraints that would prevent them from exploiting the full benefits of style rotation. Therefore, in our sample, we consider that funds experiencing style rotation alter their risk characteristics the most, and consequently, we classify them as funds with the highest level of risk-shifting.

[12] See for instance Levis and Liodakis (1999) and Clare et. al (2010) for the UK evidence.

2) ***Style drifting funds:*** those whose beta goes from significant (positive or negative) to insignificant – or vice versa – before and after the regime change. We place those funds in the moderate risk-shifting category and acknowledge that this type of change can be a result of a natural drift in a fund (e.g. stocks in a small-cap fund gradually getting larger and as a result the fund appears to be drifting towards mid-cap category, while no actual change in holdings has occurred). However, note that those natural changes will be affecting factor loadings of the fund's benchmark as well, so by applying benchmark-adjusted model to identify structural changes in style betas, we isolate those changes that are the result of true drift by fund managers beyond that of the benchmark.

3) ***Style strengthening/weakening funds:*** those whose factor loadings increase or decrease before and after the break but remain of the same sign and significance. This can happen as a result of portfolio manager's picking of more (or less) extreme stocks within the same style or naturally (e.g. if the average P/E and market value of the portfolio increases/decreases over time). Similarly, as in 2) the funds exhibiting a natural shift that is also embedded in the benchmark will be eliminated and only those whose fund managers make active increases/decreases in portfolio size or P/E ratio, for instance, will be captured. We consider this to be the least intense risk-shifting, as the fund remains within the same style classification in both regimes and would agree with investor's initial risk preferences.

Table 2 shows the number of risk changes that result in style rotation (darkest shade), style drifting (lightest shade) and style strengthening/weakening (medium shade, diagonal of the matrix) from period t to t+1, i.e. before and after *each* structural break.

Table 2: The number of breakpoints in factor loadings, per style and intensity of style change

The Table is organised as a matrix and shows number of shifts from t to t+1 for each style combination: For instance, first cell in the first row shows that there were 19 shifts when a fund was large value in period t (before the break) and strengthened/weakened their style but remained in large-value category in period t+1 (after the break); there are 22 shifts when a fund was large value in t and moved to large-blend in t+1, etc. Last column and row show the totals.

	t+1 t	LargeValue	LargeBlend	LargeGrow	MidValue	MidBlend	MidGrowth	SmallValue	SmallBlend	SmallGrow	Total
	LargeValue	19	22	2	10	10	1	0	0	0	64
Style Rotati	LargeBlend	15	72	9	17	53	6	1	2	0	175
Style driftir	LargeGrowth	0	11	5	0	16	8	1	0	1	42
Style streng	MidValue	18	16	2	44	59	6	7	8	2	162
	MidBlend	14	70	13	50	333	67	14	45	10	616
	MidGrowth	0	5	2	1	61	38	0	18	21	146
	SmallValue	1	0	0	16	21	5	23	24	2	92
	SmallBlend	0	4	1	10	52	24	19	56	17	183
	SmallGrowth	0	0	0	1	16	16	2	11	32	78
	Total	67	200	34	149	621	171	67	164	85	1558

The table is organized as a matrix, where rows represent style of fund in period *t* (before breakpoint) and columns are style of fund in *t+1* (after breakpoint) and should be interpreted as follows. For instance, the first cell in the table shows that in 19 instances in our sample period funds that were classified as Large Value funds before the break - have strengthened or weakened their style following the break in benchmark-adjusted factor loading, remaining in the Large Value category. Then, the cell below shows that in 15 instances funds that were classified as Large Blend in period *t* changed their style to Large Value in period *t+1* following the break. And so on. It becomes evident that very few changes in style risk exposure (9 in total) result in style rotation. That is good news for investors, as style switching implies a significant shift in risk of the fund, often outside the risk parameters, the fund's official investment style corresponds to, 927 out of 1,558 style changes result in what we classify in Section 2.2.3. as style drift, while 622 changes are reflecting strengthening or weakening existing style exposure (obtained as the sum of the values in the highlighted matrix diagonal in Table 2). The table also reveals some less desirable news for investors: it shows that the fund managers seem not to apply what the literature on style investing suggests. Specifically, the vast majority style changes in Table 2 show that funds move towards the mid-cap or blend style, rather than into pure value or growth, small-cap or large-cap styles, which are proven to perform very well historically in different periods as mentioned in Section 1 of this paper.

Note that Table 2 includes all benchmark-adjusted risk shifts in our sample, even those estimated using betas based on short break periods, as explained in the previous section. Where relevant in the analysis, we will test validity of our results by removing risk shifts based on breaks shorter than 24 months in an attempt to remove bias stemming from potential error in beta estimation over short periods.

2.2.4. Performance evaluation of funds that exhibit risk changes

As a final step in our study, we evaluate whether a fund's shift in style betas results in better risk-adjusted fund performance. Our performance evaluation has four aspects.

First, we examine whether funds that go through one or more structural breaks perform better than funds with no benchmark-adjusted breaks, suggested by the AGT model. We further this analysis by assessing the performance of funds with breaks vs. 'no breaks' funds within each of the nine Morningstar style categories the fund belongs to at the start of our analysis. To complete this, we compare standard indicators of mutual fund performance such as Sharpe ratio and Fama-French three-factor model alpha (FF3) as per equation (2) as well as benchmark-adjusted (AGT) alpha given by the model from equation (1). We expect that if funds with one or more structural breaks in benchmark-adjusted risks perform better, they will either have a higher return and/or lower standard deviation, resulting in better risk-adjusted performance than funds with consistent risk exposures.

Second, we evaluate whether any particular type of risk-shifting as described in Section 2.2.3 (style rotation, style drift, or style strengthening/weakening) contributes to greater improvement in performance in the period following the break-in risk structure.

Third, we consider whether performance of those funds that exhibit risk shifting is driven by shifting their exposure to any particular investment style(s) (i.e. are funds that perform best those that move into for instance Small Value category following the structural break?). To this end, we determine the differences in the above-mentioned performance measures pre and post regime change for each style combination, as per Table 2. If a fund was successful in changing its risk structure between period t and $t+1$, it should result in an improvement in performance indicators in period $t+1$. Presenting those results using the format as per Table 2 allows us to identify if style rotation, style drift, or style strengthening or weakening have more pronounced changes in performance.

Finally, we consider whether the top 10% of funds that generate the highest excess returns in our sample perform distinctly better than the funds in the bottom decile on a risk-adjusted basis and if so, is there anything they do differently in terms of risk shifting to bottom decile funds that may contribute to their superior performance?

3. Empirical results

3.1 Performance of funds with no benchmark-adjusted structural breaks vs. funds with breaks

Let us first compare the performance of funds that do not make significant shifts in risk, identified by our benchmark-adjusted structural break analysis, to those that do so. Table 3 shows panel average annualised excess return (return of the fund minus US 3-month T-bill), standard deviation, Sharpe ratio, FF3 alpha, and AGT alpha for funds with no breaks in benchmark-adjusted style betas on one hand and funds with 1, 2, 3, 4 or 5 breaks in betas in our sample. The final row represents average performance indicators for *all* funds with significant benchmark-adjusted style and/or size risk changes in our sample. The performance indicators reported correspond to the whole sample period.

Table 3: No breaks funds vs. funds with structural breaks: performance indicators
The table shows the panel average excess return, standard deviation, Sharpe ratio, FF3 alpha, and AGT alpha for funds with no structural breaks and funds with 1,2,3,4, and 5 structural breaks as well as all funds with breaks. All values are annualised.*, **, *** denote whether funds with breaks have significantly different performance indicators from funds with no breaks at 10%, 5% and 1% respectively, based on a z-test

Funds with….	Number of funds (breaks)	Excess Return p.a. (%)	Std. Dev. p.a. (%)	Sharpe Ratio p.a.	FF3 Alpha p.a. (%)	AGT Alpha p.a. (%)
no breaks	914 (0)	5.18	18.84	0.26	-0.84	-0.94
1 break	55 (55)	5.54*	17.36	0.3	-0.17*	-0.27*
2 breaks	12 (24)	7.71***	17.55	0.53***	0.58*	0.40*
3 Breaks	7 (21)	7.24***	16.96	0.49	0.15	0.13
4 Breaks	7 (28)	5.40	15.20	0.4	-0.58	-0.70
5 Breaks	286 (1,430)	5.77***	18.42	0.28	-0.26***	-0.36***
All funds with breaks	367 (1,558)	5.81***	18.16	0.30*	-0.22***	-0.33***

Before analysing this table in more detail, we recall that the AGT model used in breakpoint analysis in this paper identifies the funds that make changes in style and size risk of their portfolios in excess of the changes embedded in the benchmark index (S&P 500). With this in mind, our results point out that funds that preserve the factor risks over time or at least keep the changes in their factor risks in line with those of the benchmark, do not perform as well as those that deviate more from the changes embedded in the index. All funds with at least one significant shift in risk, different from that of the benchmark, generate higher excess returns, lower standard deviations, and improve risk-adjusted performance as measured by Sharpe ratio, FF3 alpha, and AGT alpha. Each group of funds with benchmark-adjusted breaks generates higher Sharpe ratios and improves upon FF3 and AGT alphas relative to no-breaks group. The differences in these performance indicators between no-breaks funds and all funds with breaks are significant at 1% level (excess returns, FF3 and AGT alpha) and 10% level (Sharpe ratio) according to the z-test. Funds with one, two and five breaks achieve significantly better FF3 and AGT alphas compared to no-breaks funds. Having said this, note that factor-risk-adjusted performance reflected in both FF3 and AGT alphas is poor overall for the funds in our sample, whether they shift risks or not, which is in line with vast evidence in the literature on mutual fund performance. This does not mean, however, that there are no funds in the sample exhibit significantly positive alphas[13]. In our sample, funds with two and three benchmark-adjusted breaks realise positive FF3 and AGT alphas, which are economically different from alphas of funds with no breaks (a difference of 1.42% p.a. and 1.34% p.a. respectively). Alphas of funds with two breaks are also statistically different from alphas of funds with no breaks. Table 3 also indicates that the group of funds whose performance is most statistically significantly distinct from funds with no changes in risk are the funds with five risk shifts. Broadly speaking, these results are at odds with some studies that use holdings-based approach to determine changes in risks and do not account for the natural risk changes embedded in the funds' benchmark, such as Huang et al. (2011) who find that risk-shifting leads to poorer performance of US mutual funds. On the other hand, our results are in line with Andreu et al. (2019) who investigate holdings-based risk shifting in Spanish mutual funds. In addition, our findings concur with Wermers (2012) study that uses a detailed holdings approach and accounts for both active and passive (natural) drifts and finds that investing in funds that are style consistent is not as rewarding for investors as suggested by Brown and Harlow (2002) or Huang et al. (2011). Wermers (2012) draws attention that particular investment styles drift more and may benefit more from style drifts than others, which we will investigate in the following parts of our paper.

[13] See for instance Cuthbertson et al. (2010) and Mateus et al. (2019a) for review of literature on mutual fund performance.

3.1.1. Sub-sample Analysis

Figure 1 in Section 2.2.1. illustrated that the highest number of breaks in betas in our sample occurred in 2007, before the start of the financial crisis. Glode (2011) and Kosowski (2011) argue that good funds managers will aim to achieve outperformance during market and economic downturns, when it matters most to investors. To compare the performance of funds with and without structural breaks in betas during the known crisis periods - the dot com bubble burst in early 2000s and the financial crisis of 2008/09 - we split the period of analysis in five sub-samples. Period 1 includes the dot-com bubble period and spans from the start of our sample January 1992 to the end of 1999; Period 2 is the dot-com bubble burst period from January 2000 to December 2002; Period 3 is covering the pre-financial crisis period including the year with most breaks (year 2007, 124 breaks) in our sample and spans January 2003-December 2007; Period 4 includes the financial crisis starting from January 2008 to December 2009, and Period 5 starts in January 2010 and lasts until the end of the sample, December 2016.

Table 4 shows annualised performance indicators (obtained through panel estimation) of all funds with breaks in betas[14] versus funds with no structural breaks, over the five sub-sample periods. During the non-crisis periods, periods 1, 3, and 5, the funds with breaks exhibit significantly better excess returns, lower standard deviations, higher Sharpe ratios (albeit not always significantly) and significant improvements (in two out of three periods) in both FF3 and AGT alphas compared to funds that are more risk consistent – corroborating findings from Table 3. For completeness, we report Israelsen (2005) modified Sharpe ratio[15] applicable when excess returns are negative, making comparison across Sharpe ratios difficult; the modification does not affect our conclusions. During the crisis periods, period 2 and 4, our results are consistent with the argument of Glode (2011) and Kosowski (2011) - all funds in our sample improve factor-risk-adjusted performance and generate positive alphas, in spite of generating negative excess returns. During the dot-com crisis period, funds that break generate positive alphas which are 0.62% (FF3) and 0.66% (AGT) p.a. higher than those of funds with no change in betas, albeit difference is not statistically significant. Similarly, during the financial crisis, we observe positive, but not significantly different alphas of both groups of funds, with and without breaks in risks. This implies that although funds prone to shifts in factor betas generate

[14] To save space, we report the results for all funds with breaks only, the results split by the number of breaks are available on request.

[15] Israelsen (2005) Sharpe $= \frac{ER}{SD^{(ER/absER)}}$

better performance in the crisis periods than in the normal or up market periods, they do not manage to significantly differ from funds that do not exhibit breaks in risks. However, in the expansionary periods when, in general, mutual fund performance is evidenced to be poor (Koslowski, 2011), the funds that engage in risk shifting generate better (i.e. more positive in terms of excess returns and Sharpe ratios and less negative in terms of alphas) performance.

Table 4: Sub Sample analysis of performance of funds with no breaks in risk vs funds with breaks

Table shows annualised excess return, standard deviation, Sharpe ratio, Israelsen (2005) Sharpe ratio, FF3 alpha and AGT alpha of funds with no breaks and all funds with structural breaks in betas in five sub periods. All results are panel estimates. Period 1 is January 1992- December 2006, Period 2 is January 2007-December 2009, and Period 3 is January 2010 – December 2016. ***, **, * denotes statistically different performance indicators at 1%, 5%, and 1% level of funds with and without breaks in factor risks, based on z-test.

Funds with….	*Number of funds*	*Excess Return p.a.(%)*	*Sharpe Ratio p.a.*	*Israelsen (2005) Sharpe ratio*	*Std. Dev. p.a. (%)*	*FF3 Alpha p.a. (%)*	*AGT Alpha p.a. (%)*
Period 1: Jan 1992 – Dec 1999							
…no breaks	*516*	13.94	1.13	1.16	17.52	-1.44	-1.75***
All funds with breaks	*247*	16.14***	1.37*	1.41**	16.72	0.10***	-0.22***
Period 2: Jan 2000 – Dec 2002							
…no breaks	*662*	-13.76	-0.60	0.005	25.24	0.58	-0.36
All funds with breaks	*290*	-13.05**	-0.55	0.002	25.44	1.30	0.30
Period 3: Jan 2003 – Dec 2007							
…no breaks	*803*	11.48	0.96	1.01	12.41	0.40	0.93
All funds with breaks	*333*	10.47***	0.97	1.00	11.09	-0.08**	0.46*
Period 4: Jan 2008 – Dec 2009							
…no breaks	*696*	-8.43	-0.30	0.1	27.22	0.95	0.90
All funds with breaks	*306*	-9.18*	-0.38	0.06	26.39	0.46	0.40
Period 5: Jan 2010 – Feb 2016							
…no breaks	*622*	9.30	0.65	0.68	15.91	-2.61	-2.79
All funds with breaks	*270*	10.43***	0.70	0.74**	15.13	-1.86***	-2.04***

Let us next examine whether funds that *shift out of* any particular investment style are improving performance relative to funds whose style remains consistent throughout the sample period.

3.2 Performance of funds with benchmark-adjusted breaks vs. funds with no breaks, by investment styles

We split the funds into nine Morningstar categories according to their investment style as follows. There are three size categories (small, medium, large-cap) and three style categories (value, blend, and growth), i.e. nine categories in total. For the funds with 'no breaks' beyond the benchmark according to the AGT model, the style is determined for the whole sample period using their FF3 factor exposures. We compare the performance indicators of those funds to the funds that start in the same Morningstar category *before* their first benchmark-adjusted break occurs. For instance, performance of funds in Large Blend 'no-breaks' category is compared to funds that start as Large Blend funds but then exhibit at least one notable change in style risks over time and move to another style category. As Table 3 demonstrates that there is a comparatively small number of funds with 1, 2, 3, and 4 breaks than with 5 breaks, splitting the analysis further by investment styles will create sub-groups with very small number of funds or no funds at all. Hence, to compare the performance of funds with breaks vis-a-vie funds with no breaks by style category, we group funds with 1- 4 breaks together. Table 5 lays out the performance comparison for funds without breaks vs. funds with 1-4 breaks, 5 breaks and all funds with breaks together. All performance indicators in the table are based on the whole sample period and estimated as a panel for each of the sub-category of fund.

Panel A of Table 5 shows the average excess returns and standard deviations per style of no-breaks funds and funds with breaks in styles for 1-4 breaks, 5 breaks and all breaks together. Panel B, C, and D in their second column show Sharpe ratios (Panel B), FF3 alphas (Panel C) and AGT alphas (Panel D) of funds without breaks resulting from panel estimations for each group of funds per investment style over the whole sample period, as in Mateus et al. (2016). The rest of the columns in Panel B, C and D show the difference in respective performance indicators between funds with benchmark-adjusted breaks and funds without breaks in style betas. Each Panel of Table 5 is split by the number of breaks (as described) and the investment style the fund started with in our sample before their first break occurred (for no breaks funds the style is determined using the whole sample).

Table 5: Performance indicators by investment style and by number of breaks: Funds with no breaks vs funds with breaks in risks

Table below shows excess returns and standard deviations (Panel A), Sharpe ratio differences (Panel B), FF3 alpha differences (Panel C) and AGT alpha differences (Panel D) of funds with no structural breaks in betas vs. funds with breaks. Funds are split by number of breaks (1-4, 5 breaks and all funds with breaks) and by investment style. Style of funds with no break is based on their FF3 coefficients over the sample period (Davies, 2001). Funds with breaks are split by styles they follow *before* the first structural break. ***, ** and * denote performance parameters associated with funds *with breaks* in betas that are *significantly different* from the corresponding parameters of funds with *no breaks* at 1%, 5% and 10% level of significance.

Panel A: Excess return (std. deviation) of funds with breaks vs no break funds, in % p.a.

Style	No breaks	1-4 Break	5 Breaks	All breaks
Large Value	5.28 (14.68)	6.15*** (14.44)	6.49*** (14.49)	6.31*** (14.46)
Large Blend	4.77 (14.98)	5.67*** (15.6)	5.46*** (15.26)	5.51*** (15.34)
Large Growth	3.69 (16.18)	6.21*** (14.4)	5.28*** (16.96)	5.53*** (16.31)
Mid Value	5.59 (17.58)	7.11*** (17.48)	5.83 (16.47)	6.23*** (16.79)
Mid Blend	3.87 (17.53)	5.08*** (15.51)	4.17 (17.41)	4.32** (17.12)
Mid Growth	2.68 (20.33)	4.78*** (20.72)	5.5*** (21.02)	5.35*** (20.96)
Small Value	6.97 (19.85)	6.56 (14.31)	7.47 (17.91)	7.37 (17.57)
Small Blend	5.5 (25.01)	5.34 (20.62)	6.64*** (20.65)	6.57*** (20.65)
Small Growth	6.77 (25.27)	6.36 (22.47)	7.56*** (22.68)	7.26 (22.62)

Panel B: Sharpe ratio *differences:* funds with breaks vs. no-breaks funds, p.a.

Style	No-breaks funds Sharpe ratio p.a.	***Difference*** **in Sharpe ratio of funds with and without breaks**		
		1-4 Break	**5 Breaks**	**All breaks**
Large Value	0.33	0.06	0.14***	0.09*
Large Blend	0.33	-0.03	0.01	0.00
Large Growth	0.19	0.23***	0.13	0.16**
Mid Value	0.31	0.14***	0.01	0.05
Mid Blend	0.21	0.11	-0.04	-0.01
Mid Growth	0.11	0.15	0.14***	0.15*
Small Value	0.37	0.21	0.04	0.06
Small Blend	0.19	0.09	0.15***	0.14***
Small Growth	0.25	0.11	0.11***	0.11*

Panel C: Panel FF3 Alpha *differences*: funds with breaks vs. no-breaks funds, in % p.a.				
Style	**No breaks funds, FF3 alpha p.a.**	***Difference* in FF3 alphas of funds with and without breaks**		
		1-4 Break	**5 Breaks**	**All breaks**
Large Value	-0.69	1.60**	0.34	1.00***
Large Blend	-0.80	0.64	0.40*	0.47**
Large Growth	-0.46	0.33	0.28	0.29
Mid Value	-1.26	1.40*	0.81*	0.98**
Mid Blend	-1.87	1.23**	0.99***	1.01***
Mid Growth	-1.68	1.52	2.04***	2.00***
Small Value	-1.04	-1.58	0.78	0.49
Small Blend	-0.75	0.76	1.22	1.18
Small Growth	1.20	-1.20	-0.66	-0.79

Panel D: Panel AGT Alpha *differences* breaks vs. no-breaks funds, in % p.a.				
Style	**No breaks funds, AGT alpha p.a.**	**Difference in AGT alphas of funds with breaks / no breaks**		
		1-4 Breaks	**5 Breaks**	**All breaks**
Large Value	-0.80	1.61***	0.32	1.00***
Large Blend	-0.90	0.59	0.33	0.40*
Large Growth	-0.62	0.35	0.31	0.32
Mid Value	-1.38	1.43*	0.86*	1.03**
Mid Blend	-1.92	1.20**	0.96***	0.99***
Mid Growth	-1.72	1.47	1.99***	1.95***
Small Value	-1.16	-1.74*	0.83	0.52
Small Blend	-0.85	0.80	1.18	1.15
Small Growth	1.12	-1.19	-0.72	-0.83

Panel A shows that excess returns of funds that switch out of their original categories are by and large significantly improved (and standard deviation is decreased) compared to funds that do not exhibit benchmark-adjusted breaks. Panels B, C and D of Table 5 show the differences in Sharpe ratios, FF3 alphas and AGT alphas between funds with and without breaks. In particular, in Panels, B, C and D, we assess whether switching out of various style categories significantly improves the performance of funds with breaks relative to the comparative style category among the no-breaks funds. Identifying the positive and significant incremental values[16] in performance parameters presented panels B, C and D would indicate that funds from that particular style category are better off when they shift their risk and move to a different category, as their performance significantly improves. Similarly, worse or statistically

[16] Incremental values are the difference in performance parameters of funds with 1-4, 5 and All Breaks groups and the corresponding performance parameter of funds with no breaks across all style categories in Panels B, C and D in Table 5.

the same performance indicators would indicate that funds were better of remaining in their original style category. With this in mind, let us first focus on the last column of Table 5 in Panel B, C and D, corresponding to comparison of performance of all funds with breaks vs funds with no breaks. Those results reveal that, according to FF3 and AGT alphas (Panel C and D), switching out of the Mid-cap category pays the most, followed by switching out of the Large-cap category, while drifting away from Small-cap stocks regardless of the style (value, growth or blend) is not beneficial (incremental alphas are insignificant for all breaks group). This goes in line with the literature that provides evidence of small stocks outperformance (e.g. Fama and French, 1992). Reverting our attention to 1-4 and 5 breaks in Panel C and D, the incremental alphas are by and large insignificant except for funds in the mid-cap category, again indicating what most significant benefit of risk shifting is documented for funds shifting the risks out of the mid-cap category. Similarly, in Panel B, based on differences in Sharpe ratios across the number of breaks, most statistically significant increases in the Sharpe ratios are noted for funds that start in Mid cap, Blend and Growth styles, implying that funds that shift away from those styles tend to perform better. Note that Sharpe ratio is not as comprehensive indicator of performance as FF3 or AGT alpha, as it does not account for factor risk exposures of the funds.

As we do not have control over timing of the first break each of our funds exhibits, one might argue that estimation of investment styles our funds start with (before the first break occurs) – used as classification in table 5 - may be based on short time periods. For instance, the first break may occur a year after we have started our sample, implying that 12 months of data is used to estimate initial style of the fund before the first break. To validate our results, we create a restricted sample by removing all the funds with periods before the first break shorter than 24 months. This leaves us with 253 funds (43 with one break, 9 with two, 5 with 3, 6 with 4 and 190 with five breaks). Our results from this reduced sample are qualitatively the same and remain robust to this modification and are available on request.

In this section, we have documented which style category is most beneficial to move out from. In the next section, we break our analysis per each structural break and analyse 1) which style category would be most beneficial to switch into and 2) what intensity of risk shift (drift, style rotation, or style strengthening/weakening) is most beneficial for our funds that experience structural breaks.

4. Performance of funds following the benchmark-adjusted break in risk, by intensity of risk shift

Does the intensity of risk shift matter? To draw some general conclusions about the intensity of structural breaks in risk and performance, we draw Table 6, which provides a summary of results reported by intensity of risk shifting following a structural break. More specifically, Table 6 reports the percentage of cases in which funds have positive/negative and higher/lower performance indicators in each period t+1 (following a shift in risk) relative to period t (before the shift has occurred), grouped by three risk-shifting categories: style rotation, style drift and style strengthening/weakening. In total, as per Table 2, there is 9 style rotation shifts, 927 style drifts and 622 style strengthening/weakening shifts. The table shows that in the periods following any of the three degrees of structural change in factor risk (t+1) our funds generate positive excess returns and Sharpe ratios and lower their standard deviation compared to a pre-break period (t) in 2 out of 3 risk shifts. The excess returns and Sharpe ratios improve in about half of the cases across the three categories of funds. If we stop the analysis there and focus only on these performance parameters typically reported to investors, it would appear that the degree of risk-shifting does not matter and as long as the fund makes any adjustments to risk over time beyond those of the benchmark - it will generate positive performance in 2 out of 3 cases following the change in risk. However, when we account for factor-risk adjusted performance, those funds that suffer more extreme risk shifts resulting in style rotation perform marginally worse than the remaining two categories, decreasing their FF3 and AGT alphas in around 2/3 of the cases following the change. Hence, two conclusions can be drawn from Table 6: 1) more extreme benchmark-adjusted shifts in factor risk resulting in a change in investment style does not automatically mean better factor risk-adjusted performance 2) not all the funds exhibiting structural change in betas actually improve their performance following the shift. To identify what may be driving the difference in risk-adjusted performance, characteristics and behaviour of funds that exhibit structural breaks, we split the funds in each risk shifting category into deciles according to their excess returns in the overall sample period.

Table 6: Summary of performance before vs. after the break by intensity of risk shifting

The table shows the percentage of cases (out of 1558 total risk-shifts) in which funds exhibited positive/negative and higher/lower excess return Sharpe ratio, FF3 and AGT alpha, and lower/higher standard deviation in period t+1, following the shift. The results are presented by intensity of risk-shift: style rotation, style drift and style strengthening/weakening defined as per Section 2.2.3.

Performance t+1 versus t			
% of switches with….	**Style rotation (9 shifts)**	**Style drift (927 shifts)**	**Style strengthening /weakening (622 shifts)**
positive/negative excess return in t+1	67%/33%	68%/32%	69%/31%
positive/negative Sharpe ratio in t+1	67%/33%	68%/32%	69%/31%
positive/negative FF3 alpha in t+1	55%/45%	37%/63%	39%/61%
positive/negative AGT alpha in t+1	55%/45%	38%/62%	38%/62%
lower/higher std. deviation in t+1	67%/33%	52%/48%	58%/42%
higher/lower excess return in t+1	55%/45%	47%/53%	45%/55%
higher/lower Sharpe ratio in t+1	55%/45%	48%/52%	47%/53%
higher/lower FF3 alpha in t+1	33%/67%	46%/54%	47%/53%
higher/lower AGT alpha in t+1	33%/67%	46%/54%	48%/52%

Note that intensity of risk shift in each break accounted for in Table 6 is defined as the change in size/style betas in two consecutive periods as described in Section 2.2.3 of the paper. As we do not impose the minimum length of the break periods in order not to miss important break periods, some of the periods over which factor betas are estimated are short, potentially leading to an estimation error. To validate our findings, we repeat the analysis after removing all the risk shifts based where one (or both) consecutive break periods used to define it is shorter than 24 months. This results in total of 959 shifts (down from 1558): 8 style rotations, 614 drifts and 337 style strengthening or weakening. The results from the reduced sample confirm our main results and verify that our conclusions are not affected by beta estimation over shorter break periods. They are available on request.

5. Characteristics and behaviour of best performing funds with breaks: Top 10% vs. Bottom 10% of funds

In this section, we attempt to answer several questions:

1) Does structural change in risk imply similar performance for all funds in the group? In other words, do funds in the top excess returns decile significantly outperform funds in the bottom decile on a risk-adjusted basis, making us conclude that it is not enough just to shift-risk but it is important *how* we shift risks as well.
2) Are all funds similar in the number of risk shifts and the length of shift, i.e. are funds that shift risk more often performing better?
3) Do funds in the top decile employ more **drifts or style rotations for instance,** and are they more successful with a particular type of shift?
4) Do funds in the top decile shift more into small/large/value/growth category, which has historically been known to lead to outperformance, rather than switch to blend/mid-cap style?

To address the first of these questions, whether there is a notable difference between the funds with breaks in risks, we assess the risk-adjusted performance between the top 10% and bottom 10% of funds split according to their excess returns. Do they all perform similarly well, outperforming the funds with no breaks, as we have documented in Table 3, or some of them are significantly better than the others? To make the comparison, we compute Sharpe ratios, modified Sharpe ratios using Isralesen (2005), FF3 and AGT panel alphas for top 10% and bottom 10% of funds in terms of excess returns in Table 7. Top 10% of funds have 17.32% p.a. higher excess returns, 3.2% p.a. lower standard deviation and Sharpe ratio higher by 1.09 (0.7 modified). Both FF3 and AGT alphas of top 10% of funds are positive and significant at 1% while they are negative and significant for the bottom 10% of funds. The difference between FF3 alpha of the top and bottom decile group is 7.08% p.a. and it is highly significant at 1%[17], as is the difference between the benchmark-adjusted AGT alphas of top and bottom funds (6.81%, significant at 1%). All differences in performance parameters are highly economically significant as well, supporting our earlier notion that not all funds among the funds that shift their risks do so successfully.

[17] Based on z-test.

Table 7: Are all funds with breaks performing the same? Top 10% vs bottom 10% risk-adjusted performance and characteristics of funds

The table shows performance indicators p.a. (excess return, standard deviation, Sharpe ratio, Israelsen (2005) Sharpe ratio, FF3 alpha and AGT alpha) and characteristics of shifts (average number of breaks, average length of break and number of breaks/length of time series) for top 10% and bottom 10% of funds, classified by excess returns. ***, **, * denote significance at 1%, 5% and 10% level.

Performance indicators	**Top10%**	**Bottom 10%**
Excess Return p.a.	11.70%	-5.62%
Std. deviation p.a.	20.05%	23.26%
Sharpe Ratio p.a.	0.69	-0.41
Israelsen (2005) Sharpe p.a.	0.69	-0.01
FF3 Alpha p.a.	3.65%***	-3.42%***
AGT Alpha p.a.	3.48%***	-3.33%***

Characteristics of shifts		
Average number of breaks	3.9	4.2
Average length of break	30.4 months	13 months
Number of breaks/length of time series	0.032	0.065

Table 7 gives some indication of the frequency of the risk-shifting among the top and bottom decile funds. The top decile funds deploy on average fewer risk shifts (3.9 compared to 4.2 of the bottom decile group) over longer periods, painting a picture of a less volatile strategy on average. The average length of a break period for all funds with breaks in our sample is 34 months, which indicates that bottom 10% of funds make breaks significantly shorter than the average. We test this more formally in several ways by assigning a dummy variable in panel estimation of FF3 alpha of all funds with breaks to 1) funds that have highest number of breaks per time series (top 10%); 2) periods in which shortest breaks occur (bottom 10% in terms of length), and 3) periods in which breaks with shortest length per length of fund's time-series occur (bottom 10% in terms of length). In each of these variations, short breaks are accompanied by a significant decrease FF3 alphas[18], leading to the conclusion that short temporary changes in risk do not pay off. Note that our approach identifies the funds with significant changes in their style deviation relative to the benchmark, hence if they do that more frequently, we classify them as being more active in their approach. Petajisto (2013) considers more active funds those whose weights deviate from benchmark the most (not accounting for the frequency of the changes in their portfolios) find that more active funds perform better,

[18] Results available on request.

which is in line with our general findings from Table 3 and 4, that funds with breaks which we consider more active outperform those exhibiting no breaks; but that is not the case for all funds with breaks, as per Table 7. However, these results are in line with Cremers and Pareek (2016), who account for frequency of trading and find that funds trading frequently tend to underperform even if they are highly active. In the context of our study, Table 7 shows that among the funds that shift factor risks, those that perform poorly have shorter series of returns and resort to changes in style more frequently in search for better performance, albeit not successfully. This is consistent with Huang et al. (2011) who state that younger funds are more likely to exhibit more shifts in risks and perform worse. To confirm this further, we separate performance for top and bottom 10% of the funds before and after the breaks (period t versus period t+1). Table 8 reports the results of panel estimation of performance parameters in periods before each break and after each break[19].

Top 10% of funds maintain positive performance in post break periods, earning positive and significant FF3 and AGT alphas, and the same Sharpe ratios. Note that many breaks occur prior to the crisis periods so it is possible that the reason why funds break is to avoid the losses. Overall, top 10% of funds continue with good performance following the break. In contrast, bottom 10% worsen their performance in t+1 across all performance parameters, corroborating all our earlier discussion that poor performing funds are not successful in avoiding losses by shifting risks.

Table 8: Top 10% and Bottom 10% fund performance before and after the breaks

	Top 10%		**Bottom 10%**	
Performance indicators	**Periods t**	**Periods t+1**	**Periods t**	**Periods t+1**
Annualised Excess Return	12.2%	9.2%	-4.0%	-5.5%
Annualised Std. dev.	20.8%	19.0%	24.2%	20.5%
Sharpe Ratio p.a.	0.63	0.62	-0.32	-0.32
Israelsen (2005) Sharpe p.a.	0.63	0.62	-0.012	-0.011
Annualised FF3 Alpha p.a.	4.27%***	1.66%***	-3.30%***	-3.89%***
Annualised AGT Alpha p.a.	4.15%***	1.51%***	-3.24%***	-3.69%***
No. of observations	4,819	4,928	2,423	2,304

[19] Note that if a fund goes through more than one break, periods t (before the break) become periods (t+1) as the time goes by during the life of the fund. For instance, if a fund has two breaks it has three periods in its time series. To begin with the first period is t ('before the first break'), the second period is t+1 ('after the first break'), but then the second period becomes period t ('before the second break') etc.

Let us now address the second question from the start of this Section - are better performing funds more successfully applying a particular type of risk shift. To this end, we classify each period following a break in risk into style rotation, style drift, and style strengthening/weakening for the top 10% and bottom 10% of funds. Table 9 reports the number of risk-shifts in each category and the average value of performance indicators for each type of shift, for both top and bottom funds. To begin with, style rotation is negligible in both groups (1 instance). 2/3 of risk shifts that top-performing funds make are style drifts, while bottom performing funds have a more balanced split between style drift and style strengthening/weakening (45% vs 55% in favour of most modest of shifts). Top performing funds achieve the highest annualised alphas in the style drift category, while bottom performing funds obtain the worst annualised alphas in the strengthening/weakening category. The differences in both FF3 and AGT alphas between the top 10% and bottom 10% of funds, when a style drift and style strengthening occurs are statistically significant at 1%. Other performance indicators in the table agree that the top-performing funds perform better regardless of the intensity of risk shift. Overall, we can conclude that top-performing funds employ more of the more extreme shifts in risk and when they do so, they do it more successfully compared to bottom 10% of funds.

Table 9: Top 10% and Bottom 10% performance by type of style change

The Table shows performance indicators p.a. (excess returns, standard deviation, Sharpe ratio, Israelsen (2005) Sharpe ratio, FF3 alpha and AGT alpha) and number of shifts for style rotation, drift and style strengthening/weakening withing the top 10% and bottom 10% of funds, classified by excess returns. ***, **, * denote significance at 1%, 5% and 10% level.

	Top10% funds			**Bottom 10% funds**		
Performance indicators	**Rotation**	**Drift**	**Stg/Wea**	**Rotation**	**Drift**	**Stg/Wea**
No. of shifts	1	98	49	1	74	88
Excess Return p.a.	24.90%	8.08%	11.18%	6.22%	-3.72%	-7.62%
Std. deviation p.a.	11.22%	18.43%	19.93%	12.79%	18.99%	22.19%
Sharpe Ratio p.a.	1.08	0.66	0.65	-0.10	-0.35	-0.43
Israelsen (2005) Sharpe p.a.	1.08	0.66	0.65	-0.01	-0.01	-0.01
FF3 Alpha p.a.	0.44%	1.74%***	1.33%	2.10%	-3.35%***	-4.76%***
AGT Alpha p.a.	0.59%	1.67%***	1.05%	3.14%	-3.15%***	-4.57%***

Finally, we address whether significantly better performance of top 10% of funds can be linked to larger number of shifts towards more extreme styles. In other words, are funds with the lowest excess returns ‘playing too safe’ so that their risk shifts are associated mainly with shifts towards mid-cap and blend style? Table 10 illustrates the findings. We make two observations in Table 10. First, both top and bottom decile funds make most shifts towards mid-cap and blend style. We have seen in Table 5 that switching out of those styles would be more beneficial for our funds. Second, top decile funds have considerably fewer switches towards mid and blend category compared to the bottom decile. Specifically, bottom 10% of funds make around 80% of their risk shifts towards the mid-cap and blend-style, while only 20% of shifts are towards more extreme styles that have historically led to superior performance. Top 10% of funds on the other hand shift considerably more towards more extreme small cap/large cap/value/growth styles (around 40% of the cases).

Table 10: Shifts towards different size categories

The table shows the number of shifts and % of shifts towards three size categories (large, small, medium) and three style categories (value, growth, blend) following the break in benchmark-adjusted SMB and HML factor betas for top 10% and bottom 10% of funds.

	top 10% (# of shifts)	**top 10% (% of shifts)**	**bottom 10% (# of shifts)**	**Bottom 10% (% of shifts)**
Switch into Large cap	5	3.4%	17	10.4%
Switch into Small cap	60	40.5%	16	9.8%
Switch into Mid Cap	83	56.1%	130	79.8%
Switch into Value	20	13.5%	7	4.3%
Switch into Growth	36	24.3%	27	16.6%
Switch into Blend	92	62.2%	129	79.1%

Given that many of our bottom 10% of funds have short time series and in turn short break periods over which their factor betas were estimated (average in Table 7 being 13 months), to validate the results from Tables 9 and 10, we have removed all the funds with break periods shorter than 24 months from the sample and obtained a reduced sample of 222 funds. We select the top 10% and bottom 10% of funds from that reduced sample and repeat the analysis from this section. We do not report the results in the paper to save space, but they are qualitatively similar to the ones obtained here albeit the differences between the groups are less pronounced as the worst performing funds with shortest series are removed from the sample. Further, the results are robust to different top and bottom fund thresholds, namely 5% and 20%. All robustness results are available on request.

Overall, our results in this section show that the better and more successful performance of top 10% of funds can be linked to having longer break periods and breaking less, having proportionally more risk shifts of greater intensify (style drifts in particular), and to having more risk shifts towards style categories that are proven to do historically better in the literature (small/large/growth/value) compared to the bottom 10% of funds.

Conclusions

In this paper, we contribute to the literature on active style drifts, beyond those embedded in the benchmark index, in US long-only active equity mutual funds. In particular, we assess the relationship across different degrees of factor-based style change (risk shifting) within mutual funds and their risk-adjusted performance. We isolate only changes in style risk beyond those embedded in the funds' benchmark index to eliminate any 'natural' changes in style risk resulting from changing company fundamentals over time. We investigate the characteristics of shifts of the best-performing funds within the group. The paper sheds additional light on a debate on style drift vs. style consistency approaches.

We identify that 367 out 1,281 funds in our sample have at least one structural break in benchmark-adjusted SMB and/or HML factor loading values from the Angelidis et al. (2013) model over our sample period January 1992 to December 2016. Specifically, we apply Bai and Perron (2003) approach to identify structural breaks in benchmark-adjusted factor loadings over time, which at the same time denote structural changes in our fund's market, size, and style risk, altering the risk profile of the fund. Overall, 1,558 changes in risk are documented in our sample. Finally, for each structural break identified, we re-estimate the Fama-French three-factor model and obtain values of SMB and HML factor loadings before and after the change of the regime. Following this estimation, we group the funds according to three levels of intensity of risk-shifting: 1) those with the most extreme risk changes are funds whose SMB and HML betas change the sign, thus exhibiting style rotation; 2) those with medium risk change are those funds whose betas go from significant (positive or negative) to insignificant or vice versa, resulting in a (natural) style drift; and 3) those with lowest risk change, where the fund's betas increase or decrease but remain the same sign implying style strengthening or weakening.

We find that funds that exhibit changes in factor betas perform better than those that are style consistent. Funds that start as mid-cap or blend funds in our analysis improve performance the most after they change factor-risk exposure. We also document that the improvements in Fama-French three-factor model alphas and benchmark-adjusted AGT alphas occur mostly when the fund changes its risk exposure to small-cap/large-cap in terms of size and value/growth in terms of style. Risk shifting into medium size and blend portfolio does not pay off on average. Our findings reveal that only 9 out of 1,558 risk changes result in extreme risk shifting, i.e. style rotation. Even though there is empirical evidence pointing at profitability of style rotation, we do not find evidence that funds that switch styles outperform following such change. Our overall results do not conclusively support the idea that funds that change risks more do so to improve performance. Our evidence shows that improvement in alphas occurs in around half of the cases following a style drift or style strengthening weakening. This shows that not all funds that shift risks do so successfully and begs the question of whether the top-performing funds among those employing shifts in risks do something different and improve performance more than the funds in the bottom performance decile. To this end, we find that top decile funds improve risk-adjusted performance following the shift in risk significantly more than the bottom decile funds; they do have longer break periods and on average do not break as often as bottom decile funds; that they have a greater proportion of more extreme risk shifts (specifically, style drifts); and that they switch more into extreme style categories (small/large/value/growth) whereas 80% of shifts of bottom decile funds are in or within mid cap and blend categories. In conclusion, while there is no conclusive evidence that more extreme shifts in risks automatically lead to greater improvement in performance, there is evidence that successful funds will drift in style more and towards value/growth/small/large categories, and they will do so less frequently than funds that do poorly.

This study is of importance to investors aiming to in a particular sub-style of equity mutual funds as it illustrates that funds are not always style consistent and such shifts in styles are not always resulting in better performance. Our results are also of interest to fund managers who consider portfolio adjustments, as it shows that shifting towards more neutral (blend or mid-cap style) will overall not improve their portfolio performance. The study opens further avenues to explore risk shifting within particular peer-groups of funds that tend to do better (or worse) than others and investigate peer-group adjusted shifts in risks.

References

Andreu, L., Sarto, J.L., and Serrano, M., 2019, Risk Shifting Consequences Depending on Manager Characteristics, *International Review of Economics and Finance*, 62 (C), p.131-152.

Angelidis, T., Giamouridis, D. and Tessaromatis, N.,2013, Revisiting Mutual Fund Performance Evaluation, *Journal of Banking and Finance*, 37 (5), p.759-1776

Agarwal, V., Naveen D. D., and Narayan Y. N., 2011, Do hedge funds manage their reported returns? *Review of Financial Studies*, 24, p. 3281-3320.

Agarwal V., Gay G. D., and Ling, L., 2014, "Window dressing in Mutual Funds", *The Review of Financial Studies*, 27(11), p. 3133–3170.

Annaert J. and Van Campenhout G., 2007. "Time variation in mutual fund style exposures", *Review of Finance*, 11(4), p. 633-661.

Asness, C.S. Friedman, J.A., Krail R.J. and J.M. Liew., 2000, Style Timing: Value versus Growth. *The Journal of Portfolio Management*, 26(3), p.50–60.

Bai, J. and Perron, P., 2003, Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, 18(1), p. 1–22.

Bai, J. and Perron, P., 1998, Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica*, 66(1), p. 47–78.

Banz, R. W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics*, 9(1), p. 3-18

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

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

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

Brown, Harlow and Zhang, 2009, Staying the Course: The Role of Investment Style Consistency in the Performance of Mutual Funds, Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1364737

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

Chow, G. C., 1960, Tests of Equality between Sets of Coefficients in Two Linear Regressions, *Econometrica*, 28(3), p. 591–605.

Clare, A., Sapuric, S. and Todorovic, N., 2010, Quantitative or Momentum-Based Multi-Style Rotation? UK Experience. *Journal of Asset Management*, 10(6), p.370–381.

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

Cremers, K. J. M. and Pareek, A., 2016, Patient Capital Outperformance: The Investment Skill of High Active Share Managers Who Trade Infrequently, *Journal of Financial Economics*, 122 (2), p. 288-306

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

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

Davies, J. L. , 2001, Mutual Fund Performance and Manager Style, *Financial Analysts Journal*, 57 (1), p. 19-27.

Davies, J.L., Fama, E. and French, K., 2000, Characteristics, Covariances and Average Returns: 1929-1997, 55(1), p.389-406 diBartolomeo D., and Witkowski E., 1997, Mutual Fund Misclassification: Evidence based on Style Analysis, *Financial Analysts Journal*, 53, p. 32-43 .

Fama, E. F., and French, K.R., 1992, The cross-section of expected stock returns, *Journal of Finance* 47, p. 427-465.

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

Garcia, R. and Perron, P., 1996, An Analysis of the Real Interest Rate Under Regime Shifts, *Review of Economics and Statistics*, 78(1), p. 111 –125.

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

He, J., Lilian N., and Q.Wang, 2004, Quarterly trading patterns of financial institutions, *Journal of Business*, 77, p. 493-509.

Herrmann, U., Rohleder, M, and Scholz H., 2016, Does style-shifting activity predict performance? Evidence from equity mutual funds, *The Quarterly Review of Economics and Finance*, 59, p. 112-130.

Hou K, and M.A. Van Dijk, 2019, Restructuring the size effect: profitability, shocks and expected stocks returns, *The Review of Financial Studies*, 32(7), p. 2850–2889.

Huang, J., Sialm, C., and Zhang, H., 2011, Risk shifting and mutual fund performance, *Review of Financial Studies*, 24, p. 2675-2616.

Israelsen, C. (2005). A refinement of the Sharpe ratio and Information ratio, *Journal of Asset Management*, 5(6), p. 423-427.

Kim M., R. Shukla and Tomas, M., 2000, Mutual Fund Objective Misclassification, *Journal of Economics and Business*, 52, p. 309-323

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

Levis, M. and Liodakis, M., 1999, The Profitability of Style Rotation Strategies in the United Kingdom. *The Journal of Portfolio Management*, 26(1), p.73–86.

Liu, J., Wu, S. Y. and Zidek, J. V., 1997, On Segmented Multivariate Regression, *Statistica Sinica*, 7(2), p. 497–525.

Meier, I., and Schaumburg, E., 2004, Do funds window dress? Evidence for U.S. domestic equity mutual funds, *Working paper, HEC Montreal and Kellogg School of Management*.

Morey, M. R., and O'Neal, E., 2006, Window dressing in bond mutual funds, *Journal of Financial Research*, 29, p. 325-347.

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

Mateus, I.B., Mateus, C. and Todorovic, N., 2019a, Review of new trends in the literature on factor models and mutual fund performance, *International Review of Financial Analysis*, 63, p. 344–354.

Mateus I.B., Mateus C. and N. Todorovic, 2019b, Benchmark-adjusted performance of US equity mutual funds and the issue of prospectus benchmarks, *Journal of Asset Management* 20(1), p. 15–30.

Musto, D. K., 1997, Portfolio disclosure and year-end price shift, Journal of Finance 52, p. 1563-1588.

Musto, D. K., 1999, Investment decisions depend on portfolio disclosures, *Journal of Finance* 54, p. 935-952.

Ng, L., and Wang, Q., 2004, Institutional trading and the turn-of-the-year effect, *Journal of Financial Economics*, 74, p. 343-366.

Petajisto, A., 2013, Active share and mutual fund performance, *Financial Analysts Journal*, 69(4), p. 73-93

Quandt, R. E., 1958, The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes, *Journal of the American Statistical Association*, 53(284), p. 873–880.
Rapach D. E., and Wohar, M. E., 2006. Structural Breaks and Predictive regression Model of Aggregate U.S. Stock Returns, *Journal of Financial Econometrics*, 4(2), p. 238-274.
Reinganum, M. R., 1981, Misspecification of capital asset pricing: empirical anomalies based on earnings yields and market values, *Journal of Financial Economics*, 9, p.19-46.
Reinganum, M. R., 1992, A revival of the small firm effect, *Journal of Portfolio Management*, 18, p.55-62
Sensoy, B.A., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry, *Journal of Financial Economics*, 92(1), p.25-39.
Sias, R.W., and Starks, L. T., 1997, Institutions and individuals at the turn-of-the-year, *Journal of Finance*, 52, p. 1543-1562.
Van Dijk, M.A., 2013, Is size dead? A review of the size effect in equity returns. *Journal of Banking and Finance*, 35(12), pp3263-3274.
Wermers, R., 2012. A matter of style: the causes and consequences of style drift in institutional portfolios, *CFR Working Paper*. No. 12-04.