

Which stocks drive the size, value, and momentum anomalies and for how long? Evidence from a statistical leverage analysis

Kevin Aretz¹ · Marc Aretz²

Published online: 5 February 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract A large number of neoclassical, behavioral, and bias-based theories try to explain the tendency of small, value, and winner stocks to outperform big, growth, and loser stocks, three well-known characteristic anomalies. Because the theories often predict similar relationships between a stock's propensity to contribute to the anomalies and a set of correlated firm characteristics, existing studies focusing on single theories do not tell us which theory is most successful in explaining the anomalies. To fill this gap, we use a new non-parametric methodology to run a horse race between the theories. In the first step, we use statistical leverage analysis to find out which stocks are ultimately responsible for the anomalies. In the second, we use the firm characteristics suggested by the theories to forecast the identity of the anomaly drivers, with the purpose of determining which theory is most supported by the data. We find that behavioral theories are most convincing in explaining the size and book-to-market anomalies, while no theory is convincing in explaining the momentum anomaly.

Keywords Characteristic anomalies · Statistical leverage analysis · Efficient markets

JEL Classification G11 · G12 · G15

✉ Kevin Aretz
kevin.aretz@mbs.ac.uk

Marc Aretz
marc.aretz@akustik.rwth-aachen.de

¹ Accounting and Finance Division, Manchester Business School, The University of Manchester, Crawford House, Booth Street West, Manchester M15 6PB, UK

² Department of Economics and Finance, RWTH Aachen University, Templergraben 64, 52062 Aachen, Germany

1 Introduction

Prior research shows that several firm characteristics explain the cross section of stock returns even when controlling for rational asset pricing factors, such as the market beta. Premier among these firm characteristics are market capitalization (“size”), the book-to-market (“BM”) ratio, and the medium-term past (“momentum”) return (Banz 1981; Rosenberg et al. 1985; Fama and French 1992; Jegadeesh and Titman 1993). Spurred by these so-called characteristic anomalies, a large number of neoclassical, behavioral, and bias-based theories have emerged over the last years trying to explain the anomalies. While each theory finds some support in empirical tests exclusively focusing on it (or on it and a restricted set of other theories), such tests do not tell us which theory is most consistent with the data. Also, given that most firm characteristics are related, such tests do not allow us to rule out that a univariate relationship between a stock’s propensity to contribute to an anomaly and a firm characteristic is driven by the effect of another firm characteristic supporting another theory.

To address the above limitations, our article runs a comprehensive horse race between the neoclassical, behavioral, and bias-based theories. To do so, we use a new non-parametric methodology. In the first step, we apply statistical leverage analysis to identify those stocks that are most responsible for the characteristic anomalies (“anomaly drivers”). Conceptually speaking, the statistical leverage analysis looks at the change in the strength of an anomaly when excluding arbitrary subsets of stocks from our stock universe, and it chooses those stocks as anomaly drivers whose joint exclusion turns the anomaly least pronounced (Belsley et al. 1980; Davidson and MacKinnon 2004). In the second step, we use univariate and multivariate analysis to compare the identified anomaly drivers with matched stocks not contributing to the anomaly (“non-anomaly drivers”) across several firm characteristics. In these comparisons, we distinguish between anomaly drivers that would be held on the long side of a portfolio trying to exploit the anomaly (“long anomaly drivers”) and those that would be held on the short side (“short anomaly drivers”).¹ Comparing the relationships found in the data with those implied by the neoclassical, behavioral, and bias-based theories, we are able to determine which theory is most successful in explaining the characteristic anomalies.

The main advantage of our empirical design is that it allows us to determine how much the firm characteristics suggested by one theory contribute to explaining the characteristic anomalies—while controlling for a large set of other firm characteristics suggested by other theories. While portfolio formation exercises also allow us to control for other firm characteristics, it is often infeasible to go beyond three- or four-way sorted portfolios, limiting the number of other firm characteristics that we can control for. Including interactions between firm characteristics and anomaly variables,²

¹ The long anomaly drivers are the small, value, and winner stocks with abnormally high returns that drive the size, BM, and momentum effects, respectively; the short anomaly drivers are the large, growth, and loser stocks with abnormally low returns that drive the size, BM, and momentum effects, respectively.

² The anomaly variables (size, BM, and momentum) are obviously also firm characteristics. Notwithstanding, we usually refer to them as anomaly variables to distinguish them from the firm characteristics that are able to forecast the anomaly drivers according to the neoclassical, behavioral, and bias-based theories.

Fama–MacBeth (1973; FM) regressions allow us to control for a larger set of other firm characteristics. However, such regressions force us to take a parametric stance on the relationships between a stock's propensity to contribute to an anomaly and the firm characteristics, and it is not always clear that our stance is correct.³ Our empirical design is able to capture the true relationship between a stock's propensity to contribute to an anomaly and a firm characteristic independent of how the relationship looks like.

Similar to Knez and Ready (1997), we find that only 0.10–1 % of stocks are responsible for the size, BM, and momentum anomalies. The long anomaly drivers often do not have higher risk exposures than the matched non-anomaly drivers, but they tend to be more volatile, more financially distressed, more or less liquid, and more likely to be a penny stock. In contrast, the short anomaly drivers often have higher risk exposures, are more volatile, are more or less followed by financial analysts, and have more or less liquid shares.

To analyze the robustness of these relationships, we use the firm characteristics to estimate the probability of a stock becoming a long or a short anomaly driver over the next 12-month investment period. We calculate this probability using either the whole sample (in-sample) or only data available until the current month (out-of-sample). Using either set of probabilities, we show that the size and BM effects are twice as strong among stocks predicted to be anomaly drivers than among stocks predicted to be non-anomaly drivers. Digging deeper, we find that it is mostly the positive relationships between idiosyncratic volatility and distress risk, on the one hand, and the propensity of becoming a long *or* short size anomaly driver, on the other, that help us to improve on the strength of the size effect. Similarly, it is mostly a positive relationship between idiosyncratic volatility and the propensity of becoming a long *or* short BM anomaly driver that helps us to improve on the strength of the BM effect.

While some firm characteristics predict the identity of the long or short momentum anomaly drivers, they do not help us to improve on the strength of this anomaly, either in-sample or out-of-sample. Thus, the ability of these firm characteristics to condition the momentum anomaly is either weak (from an economic perspective) or unstable over time.

We also study persistence in the propensity of being an anomaly driver. We do so because many studies argue that a stock's risk characteristics evolve only slowly over time. Thus, if rational risk factors were behind the characteristic anomalies, we would expect at least some persistence in the propensity of being an anomaly driver. In contrast, if the characteristic effects were generated by behavioral bias-induced mispricing, we would expect no or little persistence if the characteristic effect were the correction of the mispricing. Alternatively, if the characteristic effect was the mispricing itself, we would expect negative persistence. Our results show that being an anomaly driver in one period fails to significantly increase the probability of becoming one in the next. Also, the long (short) anomaly drivers continue to outperform

³ An excellent example comes from Hong and Stein (1999). These authors show that micro-cap stocks do not produce a momentum anomaly, but that there is a monotonically negative relationship between market size and this anomaly among the remaining stocks. Including an interaction between market size and the momentum return in an FM regression would not capture this highly non-linear relationship.

(underperform) matched stocks for only one more investment period after the initial one.

We next turn to the question of which theory is most consistent with our findings. Both idiosyncratic volatility and distress risk can sometimes act as rational pricing factors in modern pricing theories (Merton 1987; Malkiel and Xu 2006; Li et al. 2009; George and Hwang 2010). Thus, at least at first sight, our results are consistent with systematic risk differences underlying the characteristic anomalies. However, if the anomalies were due to such differences, the long size and BM drivers should be more volatile and distressed—while the short size and BM drivers should be less volatile and distressed—than the matched stocks. Because both the long *and* short size and BM drivers are more volatile and sometimes more distressed than the matched stocks, the data do not support the rational theories.

At first sight, our results are also consistent with the possibility that market microstructure-induced biases drive the characteristic anomalies—at least if volatile and distressed stocks were illiquid and traded at low prices. However, given that we directly control for share illiquidity and share price effects, it is unlikely that market microstructure effects play a major role.

In our opinion, the most convincing interpretation is that a high idiosyncratic volatility renders the size and BM anomaly drivers difficult to arbitrage, allowing for mispricing among them. In fact, supporting Avramov et al. (2009, 2011), our evidence shows that investors seem to systematically undervalue (overvalue) small (large) distressed stocks.

The result that no existing theory is able to explain the momentum anomaly is disappointing, but consistent with this anomaly being different from others. For example, different from others, the momentum anomaly is most pronounced outside of January and in expansions (Chan et al. 1996; Chordia and Shivakumar 2002; Griffin et al. 2003; Cooper et al. 2004).

Our study contributes to a large literature developing and testing theories explaining characteristic anomalies in stock returns. One school, the neoclassical, claims that the characteristic anomalies arise because the firm characteristics capture omitted or mismeasured pricing factors (Fama and French 1992, 1993, 1995; Carhart 1997; Berk et al. 1999, etc.). Another school, the behavioral, claims that the characteristic anomalies arise because of equity mispricing. The equity mispricing persists because of limits to arbitrage, such as a high (idiosyncratic) volatility or high transaction costs (Lakonishok et al. 1994; Chan et al. 1996; La Porta 1996; Shleifer and Vishny 1997; Zhang 2006, etc.). Finally, the bias-based school claims the characteristic anomalies are spurious phenomena that are generated by data-mining or -snooping or market microstructure-induced biases. Even if the characteristic anomalies were real, this school argues that they could not be exploited due to investment restrictions or trading costs (Kaul and Nimalendran 1989; Ball et al. 1995; Lesmond et al. 2004, etc.).

Our contribution to the above literature is *not* to offer new theories trying to explain the characteristic anomalies. Instead, we recognize that there is so far no study directly comparing the validity of the testable implications generated by the existing theories. As a result, we offer a joint test of existing theories, determining which ones are relatively more and which ones are relatively less successful in explaining the characteristic anomalies.

Our study is organized as follows. Section 2 discusses the rational, behavioral, and bias-based theories trying to explain the size, BM, and momentum anomalies. It also derives testable implications from these theories. Section 3 describes the methodology used in this article. In Sect. 4, we review our proxy variables and data sources. In Sect. 5, we present our empirical results. Section 6 concludes. All technical details are given in the Appendix.

2 Hypotheses development

In this section, we look at neoclassical, behavioral, and bias-based theories trying to explain the existence of the size, BM, and momentum anomalies. In the first subsection, we review the theories. In the second, we use each theory to derive testable implications regarding the relationships between certain firm characteristics and the propensity of becoming a stock significantly contributing to an anomaly, either by producing abnormally high or low returns.

2.1 Theories explaining the characteristic anomalies

2.1.1 Neoclassical (rational expectations) theories

Neoclassical theories relying on rational expectations argue that the characteristic anomalies arise because of the differences in systematic risk between the stocks producing abnormally high returns and those producing abnormally low returns. They further claim that standard asset pricing tests do not capture these differences either due to omitted or mismeasured pricing factors. If the neoclassical theories were correct, we would always be able to transform the firm characteristics into covariances between returns and systematic factors calculated from the firm characteristics. Supporting this requirement, [Fama and French \(1993\)](#) and [Carhart \(1997\)](#) show that firm characteristic-based spread portfolios indeed explain the anomalies.⁴

Other neoclassical studies search more directly for the systematic risk factors underlying the firm characteristics. For example, [Jagannathan and Wang \(1996\)](#) and [Lewellen and Nagel \(2006\)](#) report mixed evidence about whether size and BM are efficient proxies for the conditional market beta. [Hahn and Lee \(2006\)](#), [Petkova \(2006\)](#), and [Aretz et al. \(2010\)](#) show that size, BM, and momentum are related to important macroeconomic risks. [Merton \(1987\)](#) and [Malkiel and Xu \(2006\)](#) show that, in a world in which investors are only able to invest into an investor-specific restricted set of assets, idiosyncratic volatility is positively priced and the firm characteristics could capture this pricing relationship. Under asymmetric information about firm value, the firm characteristics could also capture a positive ([Lambert et al. 2007](#)) or negative ([Johnson 2004](#)) uncertainty premium. Finally, if uninformed investors need to be compensated for asymmetric information, [Brennan and Subrahmanyam \(1996\)](#)

⁴ Unfortunately, [Daniel and Titman \(1997\)](#) show that the firm characteristics have a greater pricing ability than the exposures of spread portfolios formed using the firm characteristics.

and [Amihud \(2002\)](#) show that the firm characteristics could also capture a positive relationship between share illiquidity and stock returns.

Another possibility is that the firm characteristics capture systematic distress risk. Supporting this possibility, [Queen and Roll \(1987\)](#), [Chan and Chen \(1991\)](#), and [Fama and French \(1995\)](#) show that small and value stocks are often more distressed than big and growth stocks. Also, [Avramov et al. \(2011\)](#) show that trading strategies trying to exploit the size, BM, and momentum anomalies are often implicitly long on highly distressed stocks. Despite this, two caveats are that modern asset pricing theory does not always predict a positive distress risk premium ([Garlappi et al. 2008](#); [George and Hwang 2010](#), etc.) and that the majority of empirical studies fail to find one ([Dichev 1998](#); [Campbell et al. 2008](#), etc.).

2.1.2 Behavioral theories

Behavioral theories argue that the characteristic anomalies arise because some investors are cognitively biased and their biases create mispricing ([Lakonishok et al. 1994](#); [La Porta 1996](#); [Barberis et al. 1998](#), etc.). They further argue that more rational investors are unable to exploit the opportunities arising from this mispricing due to limits to arbitrage ([Shleifer and Vishny 1997](#)). A promising candidate for a limit to arbitrage is idiosyncratic volatility. To see this, [Wurgler and Zhuravskaya \(2002\)](#) show that the ability to hedge arbitrage risk decreases with idiosyncratic volatility. Another limit to arbitrage could be high transaction costs rendering arbitrage trades prohibitively expensive ([Xue and Zhang 2011](#)).

[Daniel et al. \(1998\)](#) propose a model in which cognitive biases become more pronounced after the receipt of good news and when there is more information uncertainty. Thus, assuming limits to arbitrage, [Cooper et al. \(2004\)](#) test whether characteristic anomalies are more pronounced in expansions (after a sequence of positive market returns), and [Zhang \(2006\)](#) tests whether they are more pronounced among stocks with more uncertain information environments. Finally, because financial distress makes stocks harder to value, [Avramov et al. \(2009\)](#) test whether characteristic anomalies are mainly driven by distressed stocks.

2.1.3 Biased-based theories

Biased-based theories argue that the characteristic anomalies are spuriously driven by academics engaging in data-mining or -snooping or by market microstructure-induced return biases ([Black 1993](#); [Kothari et al. 1995](#), etc.). The bid-ask bounce and non-synchronous trading are market microstructure biases that could be behind the anomalies. For example, [Blume and Stambaugh \(1983\)](#) show that the bid-ask bounce leads to upward bias in the returns of stocks trading at low prices. [Boguth et al. \(2011\)](#) show that non-synchronous trading leads to downward bias in the returns of value-weighted portfolios mostly invested in illiquid stocks. Other studies in this school argue that the characteristic anomalies are not really spurious, but that they cannot be exploited because of transaction costs, share illiquidity, or investment restrictions ([Lesmond et al. 2004](#)). For example, most asset managers are only allowed to invest into stocks featured in specific large stock market indexes (e.g., the Russell 1000).

2.2 Testable implications

The above theories generate testable implications regarding the relationships between certain firm characteristics and the propensity of a stock to become an anomaly driver. We summarize these testable implications in Table 1. The table does not distinguish between the anomalies because the relationships predicted by each theory do not differ across anomalies.

The neoclassical theories argue that the long anomaly drivers—those producing abnormally high returns—are riskier than otherwise identical stocks, while the short anomaly drivers—those producing abnormally low returns—are less risky. Because these theories suggest that a high market-, SMB-, HML-, or WML beta, a high idiosyncratic volatility, and a high share illiquidity signal a high systematic risk, stocks with such traits are expected to become long anomaly drivers. In contrast, because a low market-, SMB-, HML-, and WML beta, a low idiosyncratic volatility, and a low share illiquidity signal a low systematic risk, stocks with such traits are expected to become short anomaly drivers. Because neoclassical theories can produce a positive *or* negative relationship between distress risk and uncertainty, on the one hand, and systematic risk, on the other, it is impossible to predict how these firm characteristics affect the propensity of becoming an anomaly driver. However, whatever the exact

Table 1 Relationships between theories, firm characteristics, and the anomaly drivers

Firm characteristics	Capturing	Neoclassical		Behavioral		Bias-based	
		LAD	SAD	LAD	SAD	LAD	SAD
Risk exposures	Systematic risk	+	–	NA	NA	NA	NA
Idiosyncratic volatilities	Systematic risk; limits to arbitrage	+	–	+	+	NA	NA
Information uncertainty	Systematic risk; mispricing	? ^a	? ^a	+	+	NA	NA
Share illiquidity	Systematic risk; limits to arbitrage; spurious effects	+	–	+	+	+	+
Distress risk	Systematic risk; mispricing	? ^a	? ^a	+	+	NA	NA
Transaction costs	Limits to arbitrage; spurious effects	NA	NA	+	+	+	+
Bid-ask bounce	Spurious effects	NA	NA	NA	NA	+	+
Non-synchronous trading	Spurious effects	NA	NA	NA	NA	+	+
Index inclusion	Spurious effects	NA	NA	–	–	–	–

In this table, we report the relationships between several firm characteristics and the propensity of becoming an anomaly driver, as predicted by the neoclassical, behavioral, and biased-based theories meant to explain the size, BM, and momentum anomalies. The table distinguishes between the propensity of becoming a long (LAD) or a short (SAD) anomaly driver. However, because the predicted relationships do not differ across the anomalies, it does not distinguish between the anomalies. The column labeled “Firm characteristics” gives the names of the theory-implied firm characteristics. The column labeled “Possibly Capturing” shows the economic concept(s) captured by the theory-implied firm characteristics. A “+” (“–”) sign indicates that the propensity of becoming an anomaly driver increases (decreases) with the firm characteristic. “NA” indicates that the theory does not predict a relationship between the propensity of becoming a (long or short) anomaly driver and the firm characteristic, while a “?” indicates that the relationship is ambiguous

^a Although the sign of the relationship is ambiguous, the firm characteristic needs to predict the long and short anomaly drivers with opposite signs in order to support this theory

relationships are, theory predicts that they condition the propensity of becoming a long anomaly driver with the opposite sign from the propensity of becoming a short anomaly driver.⁵

Behavioral theories argue that both the long *and* short anomaly drivers are difficult to arbitrage and thus mispriced. Thus, variables positively (negatively) correlated with limits to arbitrage are expected to be positively (negatively) related to the propensity of becoming a long- *or* short anomaly driver. More specifically, because a high volatility and high transaction costs create limits to arbitrage, the long and short anomaly drivers are predicted to be associated with a high volatility and high transaction costs. Also, if financial distress renders stock valuation harder, distress risk is expected to forecast the identity of the long and short anomaly drivers with a positive sign. Finally, if cognitive biases increase with information uncertainty, both the long and short anomaly drivers are expected to suffer from high uncertainty.

Some bias-based theories claim that the characteristic anomalies are spuriously driven by market microstructure biases. Because market microstructure biases are most pronounced at low share prices and share liquidity levels, penny stocks and illiquid stocks (as, e.g., identified by a low trading volume or a high fraction of zero return days) are expected to be more likely to become long *or* short anomaly drivers. Other biased-based theories claim that the characteristic anomalies cannot be exploited due to transaction costs. Because a high fraction of zero return days and a low trading volume signal high transaction costs (Kyle 1985; Admati and Pfleiderer 1988; Lesmond et al. 1999), these traits are also expected to predict the identity of the long and short anomaly drivers. Finally, if investment restrictions contribute to the anomalies, stocks featured in large indexes are less likely to become anomaly drivers.

3 Methodology

In this section, we review our empirical design. In the first subsection, we offer an intuitive description of how we use statistical leverage analysis to identify the set of stocks most strongly contributing to the anomalies. Next, we outline how we examine the relationships between the firm characteristics suggested by the neoclassical, behavioral, and biased-based theories and the propensity of becoming an anomaly driver. In the second subsection, we offer tests verifying our empirical design. We also compare our empirical design with an alternative one.

3.1 The statistical leverage analysis

3.1.1 Identification of the anomaly drivers

We run a statistical leverage analysis on cross-sectional regressions of stock returns on firm characteristics. To see why this makes sense, consider the following statistical

⁵ For example, if distressed stocks have a high systematic risk, then safe stocks must have a low systematic risk. Under these circumstances, a high distress risk predicts a high propensity of becoming a long anomaly driver, and a low distress risk a high propensity of becoming a short anomaly driver.

model in which the expected return, $E[r_{i,t}]$, is linear in K exogenous variables:

$$E[r_{i,t}] = r_0 + x_{i,t}^{(1)} \gamma_t^{(1)} + x_{i,t}^{(2)} \gamma_t^{(2)} + \cdots + x_{i,t}^{(K)} \gamma_t^{(K)}, \quad i = 1, 2, \dots, N, \quad (1)$$

where $r_{i,t}$ is the return of stock i in month t , r_0 is the return of an asset which has zero values on all the exogenous variables, $x_{i,t}^{(k)}$ is the k th exogenous variable, $\gamma_t^{(k)}$ is the slope coefficient on the k th exogenous variable, and N is the number of stocks.

To form zero investment portfolios, we write the expected portfolio return, $E[r_{p,t}]$, as:

$$\begin{aligned} E[r_{p,t}] &= \sum_{i=1}^N w_i E[r_{i,t}] = r_0 \sum_{i=1}^N w_i + \gamma_t^{(1)} \sum_{i=1}^N w_i x_{i,t}^{(1)} + \gamma_t^{(2)} \sum_{i=1}^N w_i x_{i,t}^{(2)} \\ &\quad + \cdots + \gamma_t^{(K)} \sum_{i=1}^N w_i x_{i,t}^{(K)}, \end{aligned} \quad (2)$$

where $r_{p,t}$ is portfolio p 's return in month t and w_i are the portfolio weights. To estimate the slope coefficient of exogenous variable k , we impose on the weights the restrictions that (i) $\sum_{i=1}^N w_i = 0$ (zero investment), (ii) $\sum_{i=1}^N w_i x_{i,t}^{(j)} = 1$ if $j = k$, and (iii) $\sum_{i=1}^N w_i x_{i,t}^{(j)} = 0$ if $j \neq k$. Doing so, we ensure that $E[r_{p,t}] = \gamma_t^{(k)}$. Hence, we interpret the slope coefficient as the expected return of a zero investment portfolio with unit values on one exogenous variable and zero values on all the other exogenous variables. Because there are always an infinite number of portfolio weight-sets fulfilling these restrictions, we choose from them the one set that minimizes portfolio variance, thereby also minimizing the standard error of the slope coefficient estimate.

Fama and MacBeth (1973) and Fama (1976) show that the above set of desired portfolio weights can be derived from cross-sectional ordinary-least squares (OLS) regressions of stock returns on the exogenous variables ("Fama–MacBeth (1973) methodology"). To see how this works, collect the month t -stock returns in an $[N \times 1]$ vector \mathbf{R} and a constant plus the month t -exogenous variables in an $[N \times (K + 1)]$ matrix \mathbf{X} and run an OLS regression of \mathbf{R} on \mathbf{X} . The vector of parameter estimates from this regression, $\hat{\boldsymbol{\gamma}}$, is given by:

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R} = \sum_{i=1}^N \mathbf{W}_i r_{i,t}, \quad (3)$$

where $\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, and \mathbf{W}_i is the i th column of the \mathbf{W} matrix. The $(k + 1)$ th row of \mathbf{W} gives the desired portfolio weights for exogenous variable k , and the $(k + 1)$ th row of $\hat{\boldsymbol{\gamma}}$ gives the estimate of the month t -specific slope coefficient of the k th exogenous variable. Averaging exogenous variable k 's month t -specific slope coefficient estimates over our sample period gives us an estimate of the unconditional slope coefficient of the k th exogenous variable, which is an estimate of the expected return of a zero investment portfolio with a unit value on exogenous variable k and zero values on

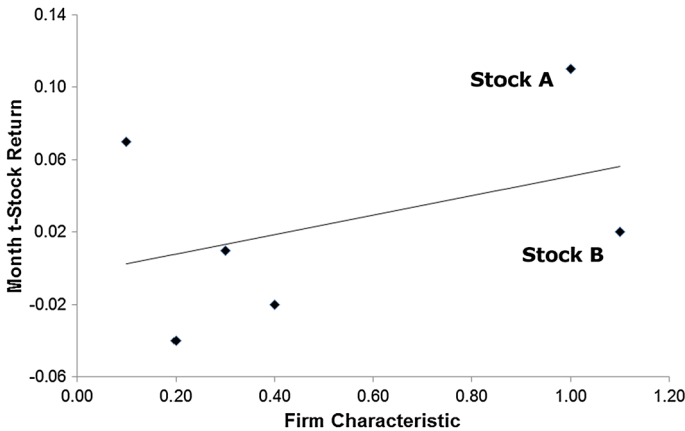


Fig. 1 Example statistical leverage analysis. This figure gives a scatter-plot showing the month t -return of six stocks on the y -axis and their values for an undefined firm characteristic on the x -axis. The figure also shows the best-fit line from an OLS regression of the returns on the firm characteristic

all others.^{6,7} In the remainder, we call the unconditional slope coefficient associated with a firm characteristic a “characteristic effect.”

Our aim is to identify the stocks that have the most positive or negative effect on a characteristic effect (“anomaly drivers”). To do so, we use statistical leverage analysis. Particularly, we define as anomaly drivers those stocks whose joint exclusion from the cross-sectional regression of \mathbf{R} on \mathbf{X} produces the largest decline in the strength of the characteristic effect (see [Belsley et al. 1980](#); [Davidson and MacKinnon 2004](#)). To see how this works, we offer an example in [Fig. 1](#). The figure is a scatter-plot showing the month t -returns of six stocks on the y -axis and their values for an undefined firm characteristic on the x -axis. The figure also gives the best-fit line (the optimal prediction) from a regression of the six stock returns on their corresponding firm characteristic-values. The slope of the best-fit line, which is around 0.053 (5.3 %), is an estimate of the month t -specific slope coefficient (effect) of the firm characteristic.

Looking at the single observation pairs, it is obvious that stock A contributes more to the 5.3 %-coefficient estimate than stock B. To be more specific, in the absence of stock A, the estimate collapses to close to zero (−0.1 %). In contrast, in the absence of stock B, the estimate almost doubles (10.1 %). The spread between the estimate

⁶ Alternatively, we are able to obtain estimates of the unconditional slope coefficients by running a single panel data regression of stock returns on firm characteristics. The parameter estimates from this regression can directly be interpreted as estimates of the unconditional slope coefficients of the K exogenous variables. Correcting for cross-sectional dependence in the residuals, [Cochrane \(2001\)](#) shows that the panel data regression is expected to produce results that are virtually identical to those from FM regressions.

⁷ In their work, [Fama and MacBeth \(1973\)](#) use portfolios as test assets in their methodology, arguing that portfolios are less subject to estimation error in their exogenous variables, especially the market beta. Because our tests do not involve the market beta, we are less worried about estimation error and thus resort to single stocks as test assets. Also, given that the CAPM was the only well-accepted asset pricing model in the early 1970s, [Fama and MacBeth \(1973\)](#) refer to only the average slope coefficient estimate on the market beta estimate as *risk* premium estimate. Whether the average slope coefficients on other exogenous variables constitute risk premia estimates has become less clear since the development of multi-factor pricing models.

excluding stock i and the full sample estimate is the statistical leverage of stock i . Thus, stock A has a statistical leverage of -5.4% ($(-0.1\%)-5.3\%$), implying that its inclusion strengthens the characteristic effect (i.e., turns it more positive). In contrast, stock B has a statistical leverage of 5.0% ($10.3-5.3\%$), implying that its inclusion dampens the characteristic effect (i.e., turns it less positive).

In our empirical tests, we determine which *set* of stocks has the most pronounced positive or negative impact on a characteristic effect while simultaneously controlling for the effects of other firm characteristics. To do so, we derive an analytical formula for the impact of excluding an arbitrary set of stocks from the regression of \mathbf{R} on \mathbf{X} in the Appendix. In theory, we could use this formula to search over all possible candidate sets until we find the desired one. However, in practice, there are too many stocks in most cross sections for this approach to be feasible. To give an example, assume there are 500 stocks in the cross section (a conservative number). We aim to identify those 25 that most strongly contribute to a characteristic effect. In this case, we would need to search over around 4.96×10^{44} possible candidate sets of 25 stocks.

Fortunately, the Appendix shows that, as the number of stocks in the cross section grows large relative to the number of excluded stocks, the effect of jointly excluding an arbitrary set of stocks converges to the sum of the effects of individually excluding the same stocks. Thus, in our empirical tests, we identify those stocks as anomaly drivers that have the most pronounced (either positive or negative) individual impacts on a characteristic effect.

In our empirical tests, we do not consider a stock's statistical leverage over a single month, but instead over the investment period from July of year t to June of year $t + 1$. However, because we assume that investors use the start of the investment period values of the firm characteristics for each month in this investment period (see Sect. 4.1), a stock's investment period-statistical leverage is simply the sum of its monthly statistical leverage estimates obtained from the cross-sectional regressions associated with the investment period. Thus, we use the sum of a stock's monthly statistical leverage estimates over this period to identify the anomaly drivers.

We split the anomaly drivers into stocks producing abnormally high returns and those producing abnormally low returns. For positively signed characteristic effects, such as the BM and momentum effects, the stocks producing abnormally high (low) returns have a negative statistical leverage and an above (below) median anomaly variable value. For negatively signed characteristic effects, such as the size effect, the stocks producing abnormally high (low) returns have a positive statistical leverage and a below (above) median anomaly variable value. Because the stocks producing abnormally high returns would be held on the long side of a portfolio trying to exploit the anomaly, we call them long anomaly drivers. Because the stocks producing abnormally low returns would be held on the short side of a portfolio trying to exploit the anomaly, we call them short anomaly drivers. The long anomaly drivers are small, value, and winners stocks; the short anomaly drivers are large, growth, and loser stocks.

To match the anomaly drivers with non-anomaly drivers, we search the 20% of stocks that contribute the least to a characteristic effect for that stock whose anomaly variable value is closest to the value of the anomaly driver. To use the size anomaly as an example, we match each size anomaly driver with that stock from the 20% weakest

contributors to the size anomaly whose size value is closest to that of the size anomaly driver.

3.1.2 Comparing anomaly- and non-anomaly drivers

Our next step is to compare those stocks that significantly contribute to a characteristic effect (the anomaly drivers) with otherwise identical firms that do not (the non-anomaly drivers). To do so, we first contrast the mean values of the theory-implied firm characteristics generated by the long or short anomaly drivers with those generated by the matched stocks. To give an example, we analyze whether the long BM anomaly drivers tend to suffer from a higher or lower share illiquidity than similar BM value-non BM anomaly drivers. To control for correlation between the firm characteristics, we also estimate the following LOGIT model:

$$\text{HLD}_{t,t+1} = \frac{1}{1 + e^{-X_t'\beta}} + \epsilon_{t,t+1}, \quad (4)$$

where $\text{HLD}_{t,t+1}$ is a dummy equal to one if a stock is classified as a size, BM, or momentum driver over the investment period and zero otherwise, X a vector containing the firm characteristics and controls measured at the start of the investment period, and $\epsilon_{t,t+1}$ the residual. To be consistent with the mean comparisons, we run the LOGIT estimations separately for stocks in anomaly decile one or ten, where the stocks in decile one (ten) are those with an anomaly variable value in the bottom (top) decile at the start of the investment period. To further control for differences in size, BM, and momentum, we include these variables as controls. To calculate unbiased inferences, we cluster standard errors at the stock-investment period level (Petersen 2009). We estimate Eq. (4) using either the entire data sample (in-sample; IS) or recursive windows of data (out-of-sample; OOS). The initial recursive window ranges from June 1974 to June 1982, and we extend the recursive window on an annual basis.

We stress that the estimates obtained from the LOGIT model in Eq. (4) do not suffer from an error-in-variables bias. While it is true that $\text{HLD}_{t,t+1}$ is estimated with error, we only use $\text{HLD}_{t,t+1}$ as endogenous variable in the LOGIT model. Thus, the estimation error inflates the volatility of the residual, but it does not bias the parameter estimates.

We use Chan et al.'s (2003) run test to study persistence in becoming an anomaly driver. The run test compares the proportions of stocks consistently classified as long or short anomaly drivers over expanding numbers of investment periods with the proportions expected under the null hypothesis of no persistence. We give a technical overview of this test—and the derivation of a test statistic showing whether the null hypothesis of no persistence can be rejected—in the Appendix. To only compare similar stocks, we conduct the run tests separately for stocks in the top or bottom size, BM, or momentum decile, where we determine inclusion in a decile using anomaly variable values measured at the start of the investment period.

We also study whether stocks classified as anomaly drivers in one investment period continue to produce abnormal returns in later investment periods. To do so, we calculate the return spread between anomaly drivers and matched non-anomaly drivers over

various post-holding periods. To adjust for other characteristic effects, we use raw *and* adjusted returns in these tests. The adjusted return is the raw return minus the return of the three-way sorted size, BM, and momentum portfolio to which a stock belongs (Daniel et al. 1997).

3.2 Verification and comparison tests

Table 2 verifies that it is reasonable to approximate the joint effect from excluding a subset of stocks with the sum of the individual effects (see Sect. 3.1.1). To achieve this goal, the table uses *real* cross-sectional data featuring all US stocks at the end of December 1986, December 1995, or December 2006. We aim to exclude from these cross sections ten, 100, 500, or 1000 stocks (# Excl.). To do so, we create one million random sets of excluded stocks for each cross section-number of excluded stocks pair. For each random set, we calculate the sum of the individual effects and the joint effect from excluding the random set from the regression of returns on size (Panel A), BM (Panel B), or momentum (Panel C).⁸ To test for bias, we regress the joint effect on the sum of the individual effects. Moreover, we calculate the Euclidean distance between the sum of the individual effects and the joint effect. Finally, we compute the ranking orders for both, subtract these from one another, take the absolute value, and sum up the absolute values ('RD'). A greater RD value indicates greater disagreement between the ranking orders obtained from the sum of the individual effects and the joint effect.

The regression constants in Table 2 suggest that there is never any constant bias in the sum of the individual effects relative to the joint effect. Also, when the number of excluded stocks is low, the slope coefficient values are all close to unity, suggesting that there is no variable bias either. However, as we exclude a larger number of stocks, the slope coefficients rise above unity, and the Euclidian distance becomes greater than zero. Notwithstanding, the R-squareds remain above 99 % and the RD statistic stays at 0.00. Taken together, these results suggest that, when excluding 500 or 1000 stocks, the joint effect becomes an order of magnitude larger than the sum of the individual effects, but the two remain almost perfectly correlated. Overall, the evidence in Table 2 suggests that our methodology works well for our purposes.

As a next step, we compare our statistical leverage approach with another approach that can be used to filter out *important* observations from regressions. Knez and Ready (1997) use Rousseeuw's (1984) least-trimmed squares (LTS) estimator to identify what they call "influential stocks" in asset pricing tests. The LTS estimator, denoted by $\tilde{\gamma}$, is the OLS estimate from the subsample of r stocks producing the lowest sum of squares:

$$\tilde{\gamma} = \left\{ (\hat{\gamma})_{\Phi} \mid \min_{\Phi} \sum_{i=\phi_1}^{\phi_r} (\hat{u}_i^2) \right\}, \quad (5)$$

⁸ We shall describe our data sources and variable definitions in greater detail below.

Table 2 Accuracy of the statistical leverage measure approximation

# Excl.	Dec-1986				Dec-1995				Dec-2006					
	<i>a</i>	<i>b</i>	R^2	RD	$\ \cdot \ $	RD	R^2	$\ \cdot \ $	RD	$\ \cdot \ $	RD	R^2	$\ \cdot \ $	RD
<i>Panel A: Size (equity market capitalization)</i>														
10	0.00	1.00	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
100	0.00	1.03	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
500	0.00	1.20	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
1000	0.00	1.50	1.000	0.01	0.00	0.00	1.000	0.01	0.00	0.01	0.00	1.000	0.01	0.00
<i>Panel B: Book-to-market</i>														
10	0.00	1.00	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
100	0.00	1.03	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
500	0.00	1.20	0.999	0.01	0.00	0.00	0.999	0.01	0.00	0.01	0.00	0.999	0.01	0.00
1000	0.00	1.50	0.998	0.03	0.00	0.00	0.998	0.02	0.00	0.02	0.00	0.999	0.02	0.00
<i>Panel C: Momentum (-3 to -1)</i>														
10	0.00	1.00	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
100	0.00	1.03	1.000	0.00	0.00	0.00	1.000	0.00	0.00	0.00	0.00	1.000	0.00	0.00
500	0.00	1.20	0.999	0.04	0.00	0.00	0.998	0.05	0.00	0.05	0.00	0.999	0.05	0.00
1000	0.00	1.50	0.996	0.14	0.00	0.00	0.995	0.15	0.00	0.15	0.00	0.998	0.17	0.00

In the table, we compare the change in the slope coefficient from (i) jointly excluding a set of stocks from a cross-sectional regression of the stock return on an anomaly variable (i.e., $\hat{\gamma}_{k+1}^{(e)}$) with (ii) the sum of the changes from individually excluding the stocks (i.e., $\sum_{e \in \Theta} (\hat{\gamma}_{k+1}^{(e)} - \hat{\gamma}_{k+1})$). To conduct the comparisons, we use cross-sectional stock data from December 1986, December 1995, or December 2006. The number of stocks excluded from these cross sections is shown under '# Excl.'. As anomaly variable, we use size (Panel A), BM (Panel B), or momentum (Panel C). For each cross section-number of excluded stocks-pair, we generate one million random sets of excluded stocks. For each random set, we calculate the joint effect and the sum of the individual effects. Under '*a*', '*b*' and ' R^2 ', we report the intercept, slope coefficient, and *R*-squared from regressing the joint effect on the sum of the individual effects. We also report the Euclidean distance between the two ($\| \cdot \|$), and the sum of the absolute differences between the rank coefficients obtained from the joint effect and the sum of the individual effects

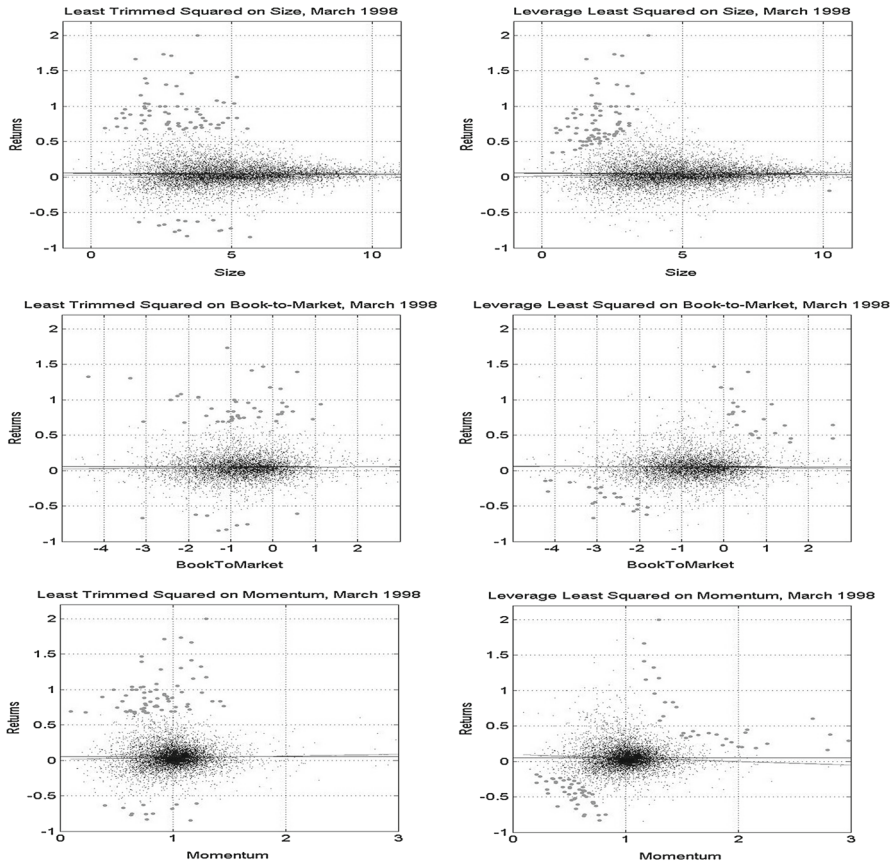


Fig. 2 Comparison statistical leverage method with least-trimmed square method. This figure shows scatter plots for the March 1998 cross-sectional regressions of the stock return on size (*upper graphs*), BM (*middle graphs*), or momentum (*lower graphs*). The *black lines* are the fitted values calculated from the full sample regression, the *gray lines* those from subsample regressions. The subsamples exclude either the 1 % of stocks whose exclusion maximizes the subsamples' R-squareds (*left graphs*) or the strongest 1 % anomaly drivers (i.e., those stocks whose exclusion produces the greatest decline in the characteristic effect; *right graphs*). The *fat dots* indicate the excluded stocks

where $r = N(1 - \alpha)$, α is the fraction of excluded stocks, N is the total number of stocks, and $\Phi = \{\phi_1, \phi_2, \dots, \phi_r\}$ is a specific subset of r stocks from a set of N stocks.⁹

To demonstrate that their influential stocks are distinct from our anomaly drivers, Fig. 2 offers scatter-plots for the March 1998-cross-sectional regressions of stock returns on size (upper panels), BM (middle panels), or momentum (lower panels). In each sub-panel, the black lines are the best-fit lines from full sample regressions. The gray lines are the best-fit lines from regressions excluding the 1 % of stocks whose

⁹ Because the computation of the LTS estimator is complicated for large samples, the estimator is usually calculated using the iterative approximation proposed by Rousseeuw and Van Driessen (2006).

exclusion maximizes the subsamples' R-squareds (left panels) and the 1 % of stocks that most strongly contribute to a characteristic effect (right panels). Excluded stocks are shown in bold. The figure shows that the two approaches exclude different sets of stocks. In particular, the statistical leverage analysis approach excludes those stocks that most strongly affect the regressions' slope coefficients, whereas the LTS-based approach excludes those stocks that produce the largest absolute residuals.¹⁰

4 Proxy variables and data

4.1 Proxy variables

Size is the natural log of the number of shares outstanding times the stock price. The BM ratio is the natural log of the ratio of the book value per share to the market value.¹¹ Momentum is the compounded return over the previous 3 months. We have chosen to study the 3-month—instead of the more commonly used 12-month—compounded return because this choice generates a stronger momentum effect in our data. Regarding timing conventions, we assume that investors observe the current size and momentum values and the 6-month lagged BM ratio values in June of each year t . They rely on these values until June of year $t + 1$, at which point they update them. Doing so, we ensure that we use only information available to investors at the time. We rely on the same conventions when forming portfolios.

We estimate a stock's market-(MKT), SMB-, HML-, and WML betas using stock-specific time-series regressions of the return on these pricing factors. We run these time-series regressions over the former 48 months of monthly data. As an alternative, we follow [Lewellen and Nagel \(2006\)](#) and estimate the market beta (MKT BETA (ALT)) by running stock-specific time-series regressions of the return on the excess market return, the 1 day-lagged excess market return, and the sum of the 2 day-, 3 day-, and 4 day-lagged excess market returns:

$$r_{i,t} = \alpha_i + \beta_{i,1}r_{mkt,t} + \beta_{i,2}r_{mkt,t-1} + \beta_{i,3}(r_{mkt,t-2} + r_{mkt,t-3} + r_{mkt,t-4})/3 + \epsilon_{i,t}, \quad (6)$$

where $r_{i,t}$ is stock i 's return over day t , $r_{mkt,t}$ is the excess market return, α_i , $\beta_{i,1}$, $\beta_{i,2}$, and $\beta_{i,3}$ are parameters, and $\epsilon_{i,t}$ is the residual. We run regression (6) over daily data from month t , and calculate the month t -market beta estimate as the sum of the slope coefficients. We include the lagged market returns in the regression to alleviate

¹⁰ We have also compared a stock's statistical leverage for the size, BM, and momentum anomalies with its exposures to the SMB, HML, and WML spread portfolios. The spread portfolio exposures are often interpreted as indicating the co-movement between a stock's return and the return of a certain class of stocks. For example, a high SMB exposure indicates that a stock's return co-moves strongly with the return of small stocks. However, despite this, there is nothing to suggest that the spread portfolio exposures capture a stock's contribution to a characteristic anomaly. Supporting this last argument, the correlation between a stock's contribution to the size (BM) [momentum] effect and its SMB (HML) [WML] exposure is only 0.0095 (−0.0045) [−0.0060].

¹¹ We also use a version of BM adjusted for deferred taxes, investment tax credits, and preferred shares (as in [Fama and French 1992](#)). Our conclusions are insensitive to which version we use.

non-synchronous trading biases. In an earlier version of this article, we also studied the macroeconomic exposures suggested by [Chan et al. \(1985\)](#) and [Chen et al. \(1986\)](#). However, because these never produced any statistically or economically significant evidence, we dropped them again from our analysis.

To proxy for idiosyncratic volatility, we use the annualized standard deviation of the residual from stock-specific market- or Carhart ([1997](#), FFC)-model estimations run over the previous 48 months of monthly data (IVOL(MKT) and IVOL(FFC), respectively). We measure distress risk using the return-on-assets (ROA), the dividend yield (DIVY), Merton's ([1974](#)) distance-to-default (DEFR), and the size decile to which a stock belonged 60 months before the current date (SIZEDEC). We investigate the lagged size decile to test the hypothesis that the anomaly drivers tend to be "fallen angels" ([Chan and Chen 1991](#)). We follow [Vassalou and Xing \(2004\)](#) in extracting the distance-to-default from Merton's ([1974](#)) model. To proxy for share illiquidity, we use the average ratio of the absolute return to trading volume ([Amihud 2002](#), ILLIQ) over the previous 12 months. Trading volume (VOL) is the mean log trading volume—and the fraction of zero return days (ZERORET) the number of zero return days divided by the number of non-missing return days—both calculated over the previous 12 months.

To study the importance of bid-ask bounce biases (which are especially pronounced at low share prices), we use a dummy variable equal to one if the share price is below one dollar and else zero (PRC; [Blume and Stambaugh 1983](#)). Because non-synchronous trading biases are most pronounced among illiquid stocks ([Boguth et al. 2011](#)), we use the share illiquidity proxies ILLIQ, VOL, and ZERORET to proxy for these. To measure information uncertainty, we derive the number of analysts providing an earnings forecast for the next fiscal year end over the prior 12 months (ANALYST). To proxy for investment restrictions, we use a dummy variable equal to one if a stock belongs to the S&P 1500 and zero otherwise (INDEX).

All exogenous variables are measured at the start of the investment period (June of year t), using only information that was available to investors at the time.

4.2 Data sources

Market data are from CRSP and accounting data from COMPUSTAT. We also use COMPUSTAT data to identify the stocks in the S&P 1500 index. Data on the benchmark factors are from Kenneth French's website. Analyst data are from I/B/E/S. Because many variables are unavailable before June 1974, our sample ranges from this date to December 2007.

5 Empirical results

This section presents our empirical results. In the first subsection, we study the strength and robustness of the size, BM, and momentum effects in our data. In the second, we compare the anomaly drivers with otherwise similar non-anomaly drivers along several theory-implied firm characteristics. The third subsection uses these theory-implied firm characteristics to construct subsamples of stocks in which the characteristic effects

are expected to be particularly strong or weak. The final subsection studies persistence in the anomaly drivers.

5.1 Strengths of the characteristic effects

In Table 3, we show the results from FM regressions of stock returns on size, BM, and momentum. Panel A gives the results from full sample estimations. Panels B and C give those from subsamples excluding specific subsets of stocks. The subsamples used in Panel B exclude those stocks whose removal maximizes the subsamples' R-squareds [Knez and Ready's (1997) LTS approach], whereas the subsamples used in Panel C exclude those stocks whose removal turns the characteristic effects least pronounced (our statistical leverage approach).

Panel A shows that the full sample creates strong size (-1.92% p.a., t -stat -3.58) and BM (3.96% p.a., t -stat 5.64) effects, but only weak momentum effects (5.52% p.a., t -stat 1.75).¹² Excluding 0.10 or 1% of the sample using the LTS-based approach eliminates the size effect, but amplifies the BM and momentum effects (Panel B). Thus, results again suggest that the LTS-based approach does not necessarily identify those stocks that are most responsible for the characteristic effects. In contrast, excluding 0.10% of the sample using the statistical leverage-based approach eliminates the size and momentum effects and reduces the BM effect to half of its former value (Panel C). Despite this, the BM effect continues to be statistically significant. Excluding 1% of the sample using the same approach turns all characteristic effects significant again, this time, however, with opposite signs.¹³

5.2 Comparison of anomaly- and non-anomaly drivers

Table 4 offers univariate comparisons of the anomaly drivers and the matched non-anomaly drivers across firm characteristics suggested by the neoclassical, behavioral, and bias-based theories to predict the anomaly drivers. A stock is classified as an anomaly driver if it ranks among the top percentile anomaly drivers over the July of year t to June of year $t + 1$ -investment period; the matched non-anomaly are from the sample of the 20% weakest contributors to the same anomaly over the same period (see Sect. 3.1.1). The first column of the table compares the whole set of anomaly drivers with all other stocks; the second and third separately compare the long and short anomaly drivers with matched non-anomaly drivers.¹⁴

¹² We annualize the per-month characteristic effect estimates in Table 3 by multiplying them by 12.

¹³ The panel data regression methodology produces a size effect of -2.21% p.a. (t -stat of -3.02), a BM effect of 5.06% p.a. (t -stat of 3.98) and a momentum effect of 3.85% p.a. (t -stat of 0.56 ; unreported). Setting the characteristic effects equal to zero requires us to exclude 0.15% of all stocks from the size estimation, 0.34% of all stocks from the BM estimation, and 0.01% of all stocks from the momentum estimation.

¹⁴ We only report empirical results based on anomaly drivers identified using FM regressions featuring only one anomaly variable. Neither controlling for other anomaly variables nor using panel data regressions in identifying the anomaly drivers (see footnotes 4 and 8) greatly changes our conclusions.

Table 3 Fama and MacBeth (1973) regressions

Anomaly	Exp sign	Sort out	Constant	Size	BM	Mom
Panel A: OLS regression						
Size	–		2.07***	–0.16***		
BM	+		1.65***		0.33***	
Mom	+		0.96**			0.46*
All	–/+/+		1.88***	–0.17***	0.21***	0.46**
Panel B: Least-trimmed square						
Trimmed at the 0.1 % level (mean of 5 observations per estimation excluded)						
Size	–		1.34***	–0.03		
BM	+		1.46***		0.34***	
Mom	+		0.41			0.82***
All	–/+/+		0.94*	–0.05	0.27***	0.68***
Trimmed at the 1 % level (mean of 54 observations per estimation excluded)						
Size	–		–0.17	0.21***		
BM	+		1.02***		0.39***	
Mom	+		–0.59			1.37***
All	–/+/+		–0.89*	0.21***	0.42***	1.00***
Panel C: Leverage-point elimination square						
Trimmed at the 0.1 % level (mean of 5 observations per estimation excluded)						
Size	–		1.28***	–0.02		
BM	+		1.53***		0.16***	
Mom	+		2.18***			–0.77***
All	–/+/+	Size	1.03**	–0.03	0.29***	0.54**
All	–/+/+	BM	1.68***	–0.17***	0.04	0.56**
All	–/+/+	Mom	2.82***	–0.12***	0.25***	–0.68***
Trimmed at the 1 % level (mean of 54 observations per estimation excluded)						
Size	–		–0.98***	0.39***		
BM	+		1.23***		–0.51***	
Mom	+		5.64***			–4.18***
All	–/+/+	Size	–1.32***	0.39***	0.57***	0.58***
All	–/+/+	BM	1.55***	–0.23***	–0.65***	0.76***
All	–/+/+	Mom	5.78***	–0.06	0.29***	–3.93***

The table shows the results from Fama–MacBeth (1973) regressions of the stock return on size, BM, and momentum, either separately or jointly. Under ‘exp sign’, we show the signs of the relationship expected from prior empirical work. Parameter estimates (est) are per month and in bold. ‘***’, ‘**’, and ‘*’ indicate that the parameter estimates are statistically significant at the 99, 95, and 90 % confidence levels, respectively. In Panel A, we perform the FM regressions on the full sample. In Panel B, we exclude the 0.10 or 1 % of stocks from each cross-sectional regression whose omission maximizes the subsamples’ R-squareds. In Panel C, we exclude those 0.10 or 1 % of stocks from each cross-sectional regression whose omission produces the greatest decline in a characteristic effect (if the characteristic effect is ambiguous, we indicate it under ‘sort out’). For the size effect, these are the stocks whose exclusion leads to the greatest *increase* in the size effect estimate. In contrast, for the BM and momentum effects, these are the stocks whose exclusion leads to the greatest *decrease* in the BM and momentum effect estimates. The sample period ranges from July 1974 to December 2007

Table 4 A univariate comparison of anomaly and non-anomaly drivers

Panel A: Size	All firms (694 obs)			Small firms (499 obs)			Large firms (195 obs)		
	Lev	No-lev	Diff	Lev	No-lev	Diff	Lev	No-lev	Diff
MKT BETA	1.29	1.11	0.18***	1.26	1.21	0.05	1.36	0.99	0.36***
MKT BETA (ALT)	0.86	0.92	-0.05	0.69	1.00	-0.31	1.32	0.98	0.34*
SMB BETA	1.07	0.81	0.26***	1.38	1.31	0.07	0.28	0.05	0.23***
HML BETA	0.00	0.16	-0.17***	0.13	0.07	0.06	-0.34	0.09	-0.43***
WML BETA	-0.26	-0.12	-0.13***	-0.33	-0.14	-0.19***	-0.07	-0.11	0.03
IVOL(MKT)	0.20	0.12	0.07***	0.23	0.22	0.02***	0.10	0.08	0.02***
IVOL(FFC)	0.18	0.11	0.07***	0.22	0.20	0.02***	0.09	0.07	0.02***
ROA	-0.09	0.02	-0.11***	-0.15	-0.07	-0.08***	0.07	0.07	0.00
DIVY	0.02	0.03	-0.02***	0.01	0.01	0.00	0.04	0.05	-0.01***
DEFR	0.18	0.03	0.15***	0.24	0.15	0.09***	0.01	0.01	0.01
SIZEDEC	5.07	6.73	-1.67***	3.50	2.68	0.82***	9.07	9.58	-0.51***
ZERORET	0.23	0.17	0.07***	0.29	0.29	0.00	0.10	0.12	-0.02***
ILLIQ	15.72	6.08	9.65***	21.66	23.98	-2.32**	0.53	0.87	-0.35**
VOL	8.98	9.86	-0.88***	7.92	7.31	0.61***	11.71	11.23	0.48***
PRC	0.28	0.02	0.26***	0.39	0.25	0.13***	0.00	0.00	0.00
ANALYST	0.85	1.40	-0.56***	0.05	0.03	0.02***	2.88	2.32	0.56**
INDEX	0.14	0.30	-0.16***	0.00	0.00	0.00	0.50	0.64	-0.14***

Table 4 continued

	All firms (614 obs)		Value firms (270 obs)		Growth firms (344 obs)	
	Lev	No-lev	Lev	No-lev	Lev	No-lev
MKT BETA	1.39	1.11	1.17	1.09	1.55	1.31
MKT BETA (ALT)	1.17	0.92	1.08	1.04	1.23	1.21
SMB BETA	1.29	0.81	1.29	1.17	1.29	1.13
HML BETA	0.14	0.16	0.47	0.37	-0.12	-0.06
WML BETA	-0.21	-0.12	-0.41	-0.26	-0.05	-0.10
IVOL(MKT)	0.20	0.12	0.18	0.16	0.21	0.18
IVOL(FFC)	0.19	0.12	0.17	0.15	0.20	0.17
ROA	-0.15	0.02	-0.06	-0.04	-0.21	-0.11
DIVY	0.01	0.03	0.01	0.01	0.02	0.03
DEFR	0.14	0.03	0.27	0.15	0.04	0.03
SIZEDEC	5.22	6.67	5.30	5.35	5.16	6.12
ZERORET	0.23	0.17	0.29	0.27	0.17	0.16
ILLIQ	11.11	6.18	20.15	16.29	4.02	4.76
VOL	9.45	9.84	8.09	8.37	10.52	10.42
PRC	0.13	0.02	0.25	0.16	0.04	0.07
ANALYST	0.46	1.39	0.23	0.43	0.63	1.30
INDEX	0.06	0.30	0.03	0.07	0.08	0.21
			Diff	Diff	Diff	Diff
			0.28***	0.08	0.08	0.25***
			0.25***	0.05	0.05	0.03
			0.48***	0.12	0.12	0.16*
			-0.02	0.10	0.10	-0.06
			-0.09***	-0.14**	-0.14**	0.05
			0.08***	0.02***	0.02***	0.03***
			0.07***	0.02***	0.02***	0.03***
			-0.17***	-0.02*	-0.02*	-0.10***
			-0.02***	0.00**	0.00**	-0.01***
			0.11***	0.12***	0.12***	0.01
			-1.45***	-0.06	-0.06	-0.96***
			0.06***	0.02**	0.02**	0.01
			4.93***	3.86***	3.86***	-0.74
			-0.39***	-0.28*	-0.28*	0.10
			0.11***	0.09***	0.09***	-0.03**
			-0.93***	-0.20**	-0.20**	-0.67***
			-0.24***	-0.04**	-0.04**	-0.12***

Table 4 continued

	All firms (694 obs)			Winner firms (491 obs)			Loser firms (203 obs)		
	Lev	No-lev	Diff	Lev	No-lev	Diff	Lev	No-lev	Diff
MKT BETA	1.40	1.11	0.29***	1.30	1.30	0.00	1.63	1.38	0.25***
MKT BETA (ALT)	1.34	0.91	0.43***	1.30	1.24	0.05	1.45	1.27	0.18
SMB BETA	1.52	0.80	0.71***	1.50	1.31	0.19***	1.56	1.43	0.13
HML BETA	0.02	0.16	-0.15***	0.10	0.08	0.02	-0.19	-0.02	-0.17
WML BETA	-0.24	-0.12	-0.12***	-0.32	-0.30	-0.03	-0.05	-0.09	0.04
IVOL(MKT)	0.21	0.12	0.08***	0.20	0.19	0.01**	0.21	0.19	0.02***
IVOL(FFC)	0.19	0.11	0.08***	0.19	0.18	0.01**	0.20	0.17	0.02***
ROA	-0.10	0.02	-0.12***	-0.07	-0.06	-0.02	-0.16	-0.06	-0.10***
DIVY	0.01	0.03	-0.02***	0.01	0.01	-0.01***	0.01	0.01	0.00
DEFR	0.11	0.03	0.08***	0.09	0.03	0.05***	0.18	0.15	0.03
SIZEDEC	4.87	6.74	-1.87***	4.77	5.12	-0.35**	5.13	4.94	0.19
ZERORET	0.23	0.17	0.06***	0.23	0.20	0.03***	0.22	0.23	-0.01
ILLIQ	11.30	6.12	5.18***	13.45	8.79	4.66***	6.11	9.89	-3.77***
VOL	9.30	9.86	-0.56***	8.91	9.49	-0.58***	10.23	9.52	0.71***

Table 4 continued

Panel C: Momentum	All firms (694 obs)		Winner firms (491 obs)		Loser firms (203 obs)	
	Lev	No-lev	Lev	No-lev	Lev	No-lev
PRC	0.10	0.02	0.09	0.06	0.12	0.11
ANALYST	0.49	1.41	0.43	0.55	0.62	0.65
INDEX	0.05	0.30	0.06	0.12	0.04	0.09
			Diff		Diff	Diff
			0.08***		0.03*	0.01
			-0.92***		-0.12	-0.04
			-0.24***		-0.06***	-0.04*

This table compares the stocks driving anomalies (anomaly drivers) with others along several firm characteristics. The anomalies are the size, BM, and momentum effects (Panels A, B, and C, respectively). An anomaly driver is a stock ranking among the top 1 % contributors to a characteristic anomaly over the investment period from July of year t to June of year $t + 1$. Under 'All Firms', we compare the anomaly drivers with all other firms. In the other columns, we match each anomaly driver that would be held long (the small, value, and winner stocks) or short (the large, growth, and loser stocks) in a trading strategy aimed at exploiting the characteristic effect with a non-anomaly driver. The non-anomaly driver is that stock from the bottom quintile contributors to a characteristic anomaly whose beginning of the investment period-anomaly variable value is closest to that of the anomaly driver. The average firm characteristic value for the anomaly drivers is shown under 'lev', while the average value for the other stocks is shown under 'no lev'. The difference is given under 'diff', with '***', '**', and '*' indicating statistical significance at the 99, 95, and 90 % confidence levels, respectively. The firm characteristics are: Estimates of the market (MKT) beta, the SMB beta, the HML beta, and the WML beta obtained from 4-year rolling window estimations using monthly data—or a market beta estimate obtained from 1-month rolling window estimations using daily data (MKT BETA (ALT)); the residual volatility from monthly 4-year rolling window regressions of a stock's return on the excess market return (IVOL(MKT)) or the market return, SMB, HML and WML (IVOL(FFC)); the return-on-assets (ROA), the dividend per share divided by the equity price (DIVY), a Merton (1974) distress risk estimate computed following Vassalou and Xing (2004, DEFR), the size decile to which a firm belonged 5 years ago (SIZEDEC); the percentage of zero return days (ZERORET), the average ratio of the absolute return to volume (ILLIQ), and the average log trading volume (VOL), all computed using daily data over the previous 12 months; a dummy variable equal to one if the share price is below one dollar and else zero (PRC); the number of analysts following a firm over the previous 12 months (ANALYST), and a dummy variable equal to one if a firm is contained in the Standard & Poor's 1500 and zero otherwise (INDEX). The sample period is July 1974 to December 2007

The first column shows that the whole set of anomaly drivers (including the long and short ones) is often systematically riskier, more volatile, and more financially distressed than the other stocks. Also, they suffer more strongly from share illiquidity, are more likely to trade at low prices, and are not followed by many analysts. Separately considering the long and short anomaly drivers, the second and third columns show that the *long* anomaly drivers often have a similar systematic risk than the matched non-anomaly drivers, but that they are more volatile and distressed. In comparison, the short anomaly drivers are often riskier (in terms of their market betas) and more volatile than the matched non-anomaly drivers.

In addition to these general conclusions, the long size effect drivers (the small stocks with abnormally high returns) are also more liquid, better covered by analysts, and more prone to trade at low prices than the matched stocks. In contrast, the short size effect drivers (the large stocks with abnormally low returns) are also more liquid and better covered by analysts.

The long BM and momentum effect drivers (the value and winner stocks with abnormally high returns) are less liquid, covered by fewer analysts, and less likely to be included in a broad stock market index than the matched stocks. However, they are also more likely to trade at low prices. In comparison, both the short BM and momentum effect drivers (the growth and loser stocks with abnormally low returns) are more distressed. However, only the short BM anomaly drivers are also followed by fewer analyst—while only the short momentum anomaly drivers also suffer from greater share illiquidity—than the matched stocks.

In Table 5, we show the results from full sample LOGIT estimations of $HLD_{t,t+1}$, a dummy variable equal to one if a stock is one of the top percentile anomaly drivers over the July of year t to June of year $t + 1$ -investment period and else zero, on the firm characteristics and the anomaly variables measured at the start of the investment period.¹⁵ The results reported in Panels A, B, and C are obtained from running estimations on stocks contained in either the top or the bottom size, BM, and momentum deciles, respectively.

The table suggests that, even in the presence of the anomaly variables, the firm characteristics capture a large fraction of the variation in $HLD_{t,t+1}$. For example, 10.9 % of the variation in becoming a long BM anomaly driver is attributable to the firm characteristics.

Starting with the small stocks, volatile and distressed (ROA and DEFR) stocks with low prices are significantly more likely to become a *long* size anomaly driver, while illiquid stocks (ZERORET and ILLIQ) are significantly less likely to do so. In contrast, it is distressed (DEFR) and illiquid (ILLIQ) large stocks with high market—but low HML—betas that are significantly more likely to become a *short* size anomaly driver (Panel A). Value stocks are significantly more probable to become a *long* BM anomaly driver if they are volatile and distressed (ROA and DEFR), trade at high prices, and have low WML betas. In contrast, it is volatile, lowly priced, but little followed growth stocks with high market-, HML-, and WML betas, but low SMB

¹⁵ IVOL(MKT), VOL, and SIZEDEC are all almost perfectly correlated with IVOL(FFC) or ZERORET. Thus, we exclude these variables from the LOGIT model estimations. We also exclude MKT BETA (ALT) because it is poorly measured at the stock level, and INDEX because it rarely varies across stocks.

Table 5 The propensity to become an anomaly driver

	Market		HML		WML		Idio.		ROA		Div		M(74)		%0-Ret		Illiq		Price		Analyst		R ²	
	Cons	Beta	SMB	Beta	Beta	Beta	Risk	Risk	ROA	Yield	DRisk	Days	Days	Days	Days	Days	Days	Days	Days	Cov	Cov	incl fc (%)	excl fc (%)	
Panel A: Size																								
Small stocks	7.95***	0.01	-0.06	0.07*	-0.09	2.54***	-1.14***	-1.41	0.74***	-2.86***	-0.01***	0.70***	-0.13	13.7	9.6									
Large stocks	-1.79	0.45*	0.01	-0.38**	-0.23	2.61	3.06	-4.88	2.36*	2.80	0.17*	0.00	-0.05	7.0	6.3									
Panel B: Book-to-market																								
Value stocks	-1.10	-0.06	-0.03	0.07	-0.21**	3.59***	-0.98**	-4.25	0.92***	-0.41	0.00	0.41*	-0.06	13.7	10.9									
Growth stocks	-4.26***	0.36***	-0.13**	0.13***	0.17**	3.35***	-0.43*	-1.92	0.23	0.36	-0.01	-0.80***	-0.26***	11.4	9.7									
Panel C: Momentum																								
Winner stocks	-0.50	-0.02	0.14***	-0.07	0.03	-0.63	-0.12	-6.68**	1.05***	-0.55	0.00	0.56*	-0.03	11.5	8.1									
Loser stocks	-1.96*	0.36***	0.00	-0.01	0.08	1.44	-0.50	4.96**	0.60**	3.59***	-0.02**	-0.38	-0.09	12.9	7.1									

The table shows the results of in-sample LOGIT estimations of a dummy variable equal to one for anomaly drivers and else zero, on theory-implied firm characteristics and controls. The anomalies are the size, BM, and momentum effects (Panels A, B, and C, respectively). To control for the anomaly variables, we run these estimations only on stocks in characteristic decile ten (upper row in each panel; the small, value, or winner stocks) or one (lower row; the large, growth, or loser stocks). The dummy variable is equal to one if a stock is ranked among the top 1 % contributors to a characteristic anomaly in the investment period from July of year *t* to June of year *t* + 1 and zero otherwise. The firm characteristics, including a constant (Cons), are described in the caption of Table 4. We also include the anomaly variables as controls among the firm characteristics. However, for the sake of brevity, we do not report their slope coefficient estimates. All exogenous variables are measured at the start of the investment period (June of year *t*). The table shows parameter estimates. ***, **, *, and * indicate that the parameter estimates are statistically significant at the 99, 95, and 90 % confidence levels, respectively. The R-squareds from models including or excluding the firm characteristics are shown under 'R² (incl fc)' and 'R² (excl fc)', respectively. The sample period is July 1974 to December 2007

betas that are more prone to become *short* BM anomaly drivers (Panel B). Finally, *long* momentum anomaly drivers are distressed (DEFER) winner stocks that have high SMB betas and trade at high prices. In contrast, *short* momentum anomaly drivers are distressed (DEFER) and illiquid loser stocks with high market betas and high dividends yields (Panel C).

The above results are bad news for neoclassical theories trying to explain the characteristic anomalies. First, the beta exposures often fail to forecast the identity of the anomaly drivers with the signs implied by these theories. Second, the neoclassical theories are inconsistent with the finding that *both* the long *and* short anomaly drivers are often volatile and distressed stocks. The only piece of evidence that could be consistent with these theories is that the short BM anomaly drivers are followed by only a few analysts. Regarding the size and BM anomalies, our results are more consistent with the behavioral stance that volatility limits arbitrage and leads to mispricing among distressed growth stocks with uncertain information environments. In addition, the finding that both the long and short size and BM effect drivers trade at low prices could indicate that market microstructure biases also contribute to these anomalies.

Somewhat disappointingly, no theory is able to explain the momentum anomaly. Although both the long and short momentum drivers are distressed, possibly supporting the behavioral theories, neither volatility nor transaction costs act as limits to arbitrage in their case.

5.3 Fine-tuned size, BM, and momentum strategies

We analyze whether the relationships discovered in the previous subsection help us to improve on the profitability of strategies trying to exploit the characteristic anomalies. For each anomaly, we, thus, generate two new samples. The first sample is designed to create a strong characteristic effect; thus, it contains all stocks except those in deciles one and ten that are *not* predicted to become anomaly drivers. The second sample is designed to create a weak characteristic effect; thus, it contains all stocks except those in deciles one and ten that are predicted to become anomaly drivers.¹⁶ The stocks predicted to become anomaly drivers are those for which the fitted values from the LOGIT model in Eq. (4) are above the median cross-sectional fitted value, and vice versa. We estimate the LOGIT model producing the fitted values either in-sample (IS; in this case, its coefficient values are given in Table 5) or out-of-sample (OOS). To ensure that the fitted values do not simply reflect the anomaly variables (size, BM, and momentum), we always set their slope coefficients equal to zero when calculating fitted values.

Table 6 shows the size, BM, and momentum effects separately for the whole sample and the two new samples. In addition to mean returns, the table also reports the alphas from the CAPM and the FFC model. The alphas are the intercepts from time-series regressions of the monthly FM regression-slope coefficients (the month t -conditional

¹⁶ We could have used a similar strategy to extract the anomaly- and non-anomaly drivers from the other eight characteristic deciles. However, given that the majority of anomaly drivers are contained in deciles one and ten, using this alternative strategy does not change our conclusions.

Table 6 continued

	All		In-sample (IS)		Out-of-sample (OOS)		Diff
		Diff	High lev	Low lev	High lev	Low lev	
Panel C: Momentum							
Pricing errors							
Mean return	0.448*	-0.259	0.349	0.607**	0.420	0.500*	-0.079
CAPM alpha	0.513*	-0.311	0.397	0.707***	0.469	0.606***	-0.136
FFC alpha	-0.374*	-0.286	-0.500**	-0.214	-0.433*	-0.325	-0.107
Exposures							
MKT BETA	0.151***	0.088	0.189***	0.101	0.184***	0.105	0.079
SMB BETA	-0.255***	0.052	-0.283***	-0.336***	-0.268***	-0.374***	0.107
HML BETA	0.402*	0.059	0.421	0.362*	0.414	0.386*	0.028
WML BETA	0.719***	-0.078	0.714***	0.792***	0.726***	0.782***	-0.056

In this table, we report estimates of the size (Panel A), BM (Panel B), and momentum (Panel C) effects derived from (i) all stocks ('All'), (ii) all stocks except those in characteristic deciles one and ten predicted to be non-anomaly drivers ('High lev'), and (iii) all stocks except those in characteristic deciles one and ten predicted to be anomaly drivers ('Low lev'). The differences between the characteristic effects obtained from the 'High lev' sample- and the 'Low lev' sample estimates are shown under 'Diff'. The stocks predicted to be anomaly drivers are those whose fitted values from the LOGIT estimation in Eq. (4) are above the cross-sectional median fitted value at the start of the investment period; the stocks predicted to be non-anomaly drivers are those whose fitted values are below the cross-sectional median fitted value. The LOGIT model-fitted values are obtained from either in-sample (IS) or recursive out-of-sample (OOS) estimations. However, in the creation of the fitted values, we set the slope coefficient estimates of the anomaly variables (size, BM, and momentum) equal to zero. The table reports the raw (mean return) and the risk-adjusted characteristic effect (alphas), both per month. The alphas are the intercepts from time-series regressions of the characteristic effects on the excess market return (CAPM alpha) or the excess market return, SMB, HML and WML (FFC alpha). The table also reports the slope coefficient estimates from time-series regressions of the characteristic effects on the same factors. *, **, ***, and *** indicate statistical significance at the 99, 95, and 90 % confidence levels, respectively. The recursive estimations start with the period from June 1974 to June 1982, so that reported parameter estimates are based on the July 1982–December 2007 period

characteristic effect estimates) on the relevant pricing factors. The relevant pricing factors are the excess market return for the CAPM, and the excess market return, SMB, HML, and WML for the FFC model.

The full sample size, BM, and momentum effects in the table are similar to those in Table 3.¹⁷ More importantly, the size effect is 0.042 % (IS) or 0.061 % (OOS) per month stronger among stocks predicted to produce a stronger size effect (lev) than among those predicted to produce a weaker one (no lev, Panel A). Given a -0.106 %-full sample size effect, the spreads are statistically and economically important. Also interestingly, the stocks predicted to produce a weaker size effect do not generate a statistically significant effect. The BM effect is 0.246 % (IS) or 0.182 % (OOS) per month more positive among stocks that are predicted to produce a stronger BM effect (lev) than among those predicted to produce a weaker effect (no lev, Panel B). Given a 0.381 %-full sample BM effect, the spreads are also statistically and economically important. Despite this, the stocks predicted to produce a weak BM effect still produce a statistically significant effect. Neither the spreads in the size effect nor those in the BM effect can be explained by the CAPM or the FFC model. To see this, note that the samples producing stronger or weaker size and BM effects generate virtually identical CAPM or FFC risk exposures.

Surprisingly, the momentum effect is 0.259 % (IS) or 0.079 % (OOS) per month weaker among stocks predicted to produce a stronger effect than among those predicted to produce a weaker. Thus, we conclude that the previously found relationships between the firm characteristics and the propensity of becoming a momentum driver are not very stable (Panel C).

Next, we turn to the question of which firm characteristics are responsible for our success in conditioning the size and BM effects. In doing so, we repeat the above analysis, this time, however, using different sets of LOGIT model-fitted values. To create these sets, we sort the firm characteristics into six mutually exclusive categories. The first set contains the risk exposures (SysRisk), the second the idiosyncratic volatility proxy (IVol), and the third the variables proxying for financial distress (DefRisk). The fourth set contains the variables proxying for share illiquidity (Illiq), the fifth the dummy variable signaling a share price below one dollar (Micro), and the sixth the analyst coverage-proxy (Uncertainty).¹⁸ For each set, we start with the slope coefficients obtained from the IS LOGIT- (see Table 5) or OOS LOGIT-models featuring all firm characteristics and anomaly variables. We then create the new IS and OOS LOGIT model-fitted values by setting the slope coefficients on all variables *except those on the variables in the set* equal to zero. The advantage of this strategy is that it allows us to analyze the ability of a specific set of firm characteristics to condition the size and BM effects while still controlling for correlation between these variables and those contained in other sets.

Table 7 shows the spreads in the anomaly effects across the sample of stocks that are expected to produce a strong effect according to one of the six sets of firm characteristics and the sample of stocks that are expected to produce a weak effect according to

¹⁷ They are not identical to the estimates in Table 3 because the current tests exclude stocks with incomplete data for the firm characteristics.

¹⁸ More details about which exogenous variables belong to which set are in the caption of Table 7.

Table 7 Spreads in the characteristic effects and the firm characteristics

	Spreads in characteristic effect induced through					
	SysRisk	IVol	DefRisk	Illiq	Micro	Uncertainty
Panel A: Size						
In-sample						
Both	0.019	-0.101***	-0.057***	0.057***	-0.030*	-0.087***
Long	0.014	-0.068***	-0.042***	0.067***	-0.013	-0.064***
Short	0.003	-0.031***	-0.013**	-0.038**	-0.007	-0.025**
Out-of-sample						
Both	-0.023	-0.101***	-0.038*	0.097***	-0.014	0.022
Long	-0.022	-0.068***	-0.027*	0.072***	-0.029***	0.042***
Short	0.003	-0.031***	-0.009	0.002	-0.007	-0.022*
Panel B: Book-to-market						
In-sample						
Both	0.058	0.140***	0.121**	-0.074	0.032	0.175***
Long	0.023	0.050	0.040	-0.068*	0.006	0.078***
Short	0.027	0.060	0.056	0.010	0.042	0.050
Out-of-sample						
Both	0.042	0.140***	-0.014	0.071	0.032	-0.034
Long	0.027	0.050	0.079**	0.029	0.003	-0.064*
Short	0.006	0.060	-0.099**	0.031	0.041	0.039
Panel C: Momentum						
In-sample						
Both	-0.151	-0.149	-0.066	-0.132	-0.049	0.113
Long	0.121	-0.040	0.086	0.006	0.058	0.159
Short	-0.185	-0.088	-0.102	-0.125	-0.055	-0.025
Out-of-sample						
Both	-0.123	-0.369	0.006	-0.117	0.233	-0.008
Long	0.046	-0.176	-0.003	-0.043	-0.204	0.088
Short	-0.095	-0.091	0.015	-0.068	0.124	-0.107

The table shows the spreads in the size (Panel A), BM (Panel B), and momentum (Panel C) effects across a sample containing all stocks except those in decile one and ten (both), ten (long), or one (short) predicted to be non-anomaly drivers and a sample containing all stocks except those in characteristic deciles one and ten, ten, or one predicted to be anomaly drivers. The spreads are per month. The stocks predicted to be anomaly drivers are those whose fitted values from the LOGIT estimation in Eq. (4) are above the cross-sectional median fitted value at the start of the investment period; the stocks predicted to be non-anomaly drivers are those whose fitted values are below the cross-sectional median fitted value. The LOGIT model-fitted values are obtained from either in-sample (IS) or recursive out-of-sample (OOS) estimations. In creating the fitted values, we set all slope coefficient estimates—except those on the firm characteristics in the indicated set—equal to zero. We focus on the following sets of firm characteristics: (i) Systematic risk (SysRisk): the MKT-, SMB-, HML- and WML betas; (ii) Idiosyncratic risk (IVol): IVOL(FFC); (iii) Default risk (DefRisk): ROA, DIVY and DEFR; (iv) Share illiquidity (Illiq): ZERORET and ILLIQ; (v) Microstructure issues (Micro): PRC; and (vi) Uncertainty (Uncertainty): ANALYST. The firm characteristics and anomaly variables are measured at the start of the investment period (June of year t), and they are described in the caption of Table 4. ‘***’, ‘**’ and ‘*’ indicate statistical significance at the 99, 95, and 90 % confidence levels, respectively. The recursive estimations start with the period from June 1974 to June 1982, so that reported parameter estimates are based on the July 1982 to December 2007-period

the same set. The spreads are calculated by either extracting stocks from both extreme deciles (“both”) or by only extracting stocks from decile one or ten (“long” and “short”, respectively). Independent of whether we use the IS- or OOS-fitted values, idiosyncratic volatility and distress risk have the greatest power to condition the size effect, with these variables helping us to improve on both the long and short side of a spread portfolio trying to exploit the anomaly (Panel A). In contrast, the share price and share liquidity only allow us to improve on the long side, and analyst following only allows us to improve on both sides in the IS (but not the OOS) tests.

Turning to the BM anomaly, idiosyncratic volatility, distress risk, and analyst coverage have the greatest conditioning power in the IS-tests. However, of these variables, only idiosyncratic volatility continues to successfully condition the BM anomaly in the OOS tests. Finally, regarding the momentum anomaly, we again find no evidence suggesting that any set of variables is successfully able to condition this anomaly, either in the IS- or the OOS tests.

Our finding that volatility and distress risk are most capable of identifying those stocks that drive the size anomaly, while volatility alone is most capable of identifying those that drive the BM anomaly, further supports the behavioral theories for these two anomalies. In addition, we find some mild evidence that market microstructure biases partially cause the abnormal returns of the long size anomaly drivers. As before, we find no evidence to suggest that either the neoclassical, behavioral, or biased-based theories explain the momentum anomaly.

5.4 Post-holding period performance

We finally study whether the stocks attracting abnormally high or low returns over the current investment period continue to do so in future periods. We do so because it is often assumed that a stock’s risk characteristics change only slowly over time. Thus, if the long anomaly drivers are systematically riskier than the matched non-anomaly drivers, we would not only expect them to outperform the others in the current investment period, but also in future periods. Similarly, if the short anomaly drivers are systematically less risky than the matched non-anomaly drivers, we would not only expect them to underperform the others in the current investment period, but also in future periods. In contrast, if mispricing underlies the anomalies, the abnormal performance would disappear over the near-term future if the anomaly represents the correction of the mispricing—or it would reverse if the anomaly represents the mispricing itself.

To study persistence in becoming an anomaly driver, Table 8 offers the results from Chan et al.’s (2003) run test. The table shows the average proportion of stocks that consistently rank among the top 25 or 50 % contributors to a characteristic anomaly over an expanding number of investment periods (one to five), where the averaging is done over all non-overlapping consecutive periods in our sample. We find that stocks classified as anomaly drivers are slightly more likely than others to be classified as anomaly drivers in future periods. The only exception are the 25 % top contributors to the long side of the size effect. However, despite the fact that we can usually reject the null hypothesis of no persistence, the deviations from the null hypothesis are so small that they hardly matter from an economic perspective.

Table 8 Run tests

		Number of consecutive years					Outperformance above 3rd quartile				
		Outperformance above median									
1	2	3	4	5	1	2	3	4	5		
Panel A: The long positions											
Small	0.502	0.267***	0.152***	0.088***	0.049***	0.252	0.056***	0.009***	0.002*	0.000	
Value	0.502	0.270***	0.148***	0.085***	0.056***	0.253	0.066	0.013	0.002	0.000	
Winners	0.502	0.253	0.136***	0.071***	0.038***	0.252	0.076***	0.023***	0.009***	0.001	
Exp	0.500	0.250	0.125	0.063	0.031	0.250	0.063	0.016	0.004	0.001	
Panel B: The short positions											
Large	0.501	0.257**	0.134***	0.070***	0.038***	0.252	0.074***	0.025***	0.009***	0.004***	
Growth	0.501	0.261***	0.140***	0.076***	0.042***	0.253	0.085***	0.033***	0.014***	0.007***	
Losers	0.501	0.260***	0.131	0.066	0.036*	0.252	0.078***	0.023***	0.008***	0.003***	
Exp	0.500	0.250	0.125	0.063	0.031	0.250	0.063	0.016	0.004	0.001	

The table shows the average proportion of stocks whose contribution to a characteristic effect consistently ranks among the top 25 or 50 % over an expanding number of investment periods (1–5), where the average is taken over all non-overlapping periods in the sample period. Under 'exp', we show the expected proportion of stocks under the null hypothesis of no persistence. To control for the firm characteristics, we separately perform the run tests on stocks in characteristic decile ten (the small, value, and winner stocks; Panel A) and one (the large, growth, and loser stocks; Panel B). ***, **, and * indicate that we can reject the null hypothesis of no persistence at the 99, 95, and 90 % confidence levels, respectively. The sample period ranges from July 1974 to December 2007

Table 9 The post-holding period performance of the anomaly drivers

	Size						Book-to-market						Momentum (-3 to -1)					
	Raw returns		DGTW returns		Raw returns		DGTW returns		Raw returns		DGTW returns		Raw returns		DGTW returns			
	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short		
Panel A: Top 1 % contributors to a characteristic effect																		
3 months	-0.013	-0.361***	0.049	-0.271***	-0.145	0.022	-0.146	0.105	0.138	-0.172	0.146	-0.053						
6 months	-0.047	-0.194***	-0.030	-0.124***	0.052	-0.119	0.032	-0.046	0.087	-0.220**	0.093	-0.201*						
1 year	-0.229***	-0.007	-0.077	0.035	-0.139	0.043	-0.047	0.034	-0.106*	0.068	-0.046	-0.058						
2 years	-0.235***	0.018	-0.051	0.007	-0.059	0.071	0.002	0.044	-0.099**	0.069	-0.075**	-0.006						
4 years	-0.220**	-0.014	-0.044	0.006	-0.155	0.087	-0.031	0.049	-0.166***	0.273	-0.070**	0.144						
Panel B: Top 5 % contributors to a characteristic effect																		
3 months	0.105*	-0.177***	0.126**	-0.138***	0.131***	-0.069	0.127**	0.008	0.145***	-0.198***	0.115***	-0.152***						
6 months	0.116**	-0.133***	0.109**	-0.087***	0.161***	-0.100***	0.132***	-0.012	0.095**	-0.159***	0.054	-0.119***						
1 year	-0.055	-0.018	0.030	0.006	-0.021	0.034	0.017	0.046	-0.059*	-0.033	-0.035	-0.045						
2 years	-0.136***	0.021	-0.020	0.011	-0.060*	0.039*	0.014	0.025	-0.035	0.028	-0.024	-0.005						
4 years	-0.184***	0.017	-0.007	0.007	-0.077*	0.027	0.008	0.011	-0.085***	0.041	-0.072	0.000						

The table shows the spread in returns between the strongest 1 % (Panel A) or 5 % (Panel B) contributors to a characteristic anomaly and matched non-contributors over post-holding periods ranging from 3 months to 5 years. The return spreads are compounded and annualized. We separately analyze firms that would be held long in a trading strategy aimed at exploiting the characteristic anomaly ('long') and those that would be held short ('short'). Also, we separately study 'raw returns' and returns adjusted for other characteristic effects ('DGTW returns'). We calculate the adjusted returns by subtracting from each stock's return the return of the three-way sorted size, book-to-market, and momentum portfolio to which the stock belongs. The three-way sorted portfolios are formed independently using tercile breakpoints, using June of each year as portfolio formation month and a holding period of 1 year. '***', '**', and '*' indicate statistical significance at the 99, 95, and 90 % confidence levels, respectively. The sample period is July 1974–December 2007

Table 9 shows the post-holding period performance of the strongest 1 or 5 % contributors to the size, BM, and momentum anomalies. We measure post-holding period performance using annualized compounded return spreads between the anomaly drivers and the matched non-anomaly drivers over various post-holding periods ranging from 3 months to 4 years. We find that the long (short) anomaly drivers initially continue to significantly outperform (underperform) the matched non-anomaly drivers, especially when we adjust returns for other firm characteristic effects. However, starting from 1 year after the initial holding period, the performance of the anomaly drivers and the non-anomaly drivers becomes undistinguishable from one another. In fact, looking at the raw returns, there is some tendency for performance to reverse over longer-term future horizons.

Overall, the above results are most consistent with the behavioral theories.

6 Summary and conclusion

We use a new methodology to conduct a comprehensive analysis of whether neoclassical, behavioral, or bias-based theories are most capable of explaining the size, BM, and momentum effects in stock returns. In the first step, we run a statistical leverage analysis to identify those stocks that are most responsible for the characteristic anomalies. In the second step, we compare the identified anomaly drivers with matched non-anomaly drivers along several firm characteristics that the theories predict to forecast the anomaly drivers. The purpose of these comparisons is to determine which theory is most consistent with the relationships between firm characteristics and the propensities to become a long or short anomaly driver found in the data.

Our tests suggest that a high idiosyncratic volatility and a high distress risk are the strongest indicators of stocks becoming long or short size anomaly drivers. In contrast, a high idiosyncratic volatility alone is the best indicator of stocks becoming long or short BM anomaly drivers. We also find some evidence suggesting that the small size drivers suffer from market microstructure biases. In contrast, we find no firm characteristics that consistently forecast the identity of the (long or short) momentum anomaly drivers. Finally, we show that there is little persistence in a stock's tendency to become a long or short anomaly driver.

Taken together, our evidence is most consistent with behavioral theories for the size and BM effects. In particular, the finding that both long and short anomaly drivers are volatile, sometimes distressed stocks support the behavioral hypothesis that these stocks are hard to arbitrage and thus mispriced. The low persistence in becoming an anomaly driver suggests that the anomalies capture temporary (rather than persistent) deviations from economic fundamentals.

Acknowledgments We are greatly indebted to Markus Schmid (the editor) and an anonymous referee for helping us to improve our work. We are also indebted to Behzet Cengiz, Massimo Guidolin, Peter Pope, Rüdiger von Nitzsch, as well as seminar participants at the ACATIS Value Conference in Frankfurt for their valuable and constructive comments and advice.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1: Technical details of the statistical leverage analysis

Consider the OLS regression of \mathbf{R} on \mathbf{X} , where \mathbf{R} is a $[N \times 1]$ -vector containing the endogenous variable, \mathbf{X} is a $[N \times (K + 1)]$ -matrix containing one and the K exogenous variables, and N is the number of observations. The $[(K + 1) \times 1]$ -vector of parameter estimates generated by this regression is $\hat{\boldsymbol{\gamma}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}$. We are interested in how excluding single or subsets of observations from the regression changes the parameter vector $\hat{\boldsymbol{\gamma}}$. Davidson and MacKinnon (2004) show that excluding observation i leads to the following change in $\hat{\boldsymbol{\gamma}}$:

$$\hat{\boldsymbol{\gamma}}^{(i)} - \hat{\boldsymbol{\gamma}} = -\frac{\mathbf{e}_i^T \mathbf{M}_X \mathbf{R}}{\mathbf{e}_i^T \mathbf{M}_X \mathbf{e}_i} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_X \mathbf{e}_i = -\frac{1}{1 - h_i} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i^T \hat{u}_i, \quad (7)$$

where $\hat{\boldsymbol{\gamma}}^{(i)}$ are the estimates from the regression excluding observation i , \mathbf{e}_i is an $[N \times 1]$ vector, with i th element equal to one and all other elements equal to zero, $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, and $\mathbf{M}_X = \mathbf{I} - \mathbf{P}_X$. Moreover, h_i denotes the i th diagonal element of matrix \mathbf{P}_X , while \hat{u}_i is stock i 's residual from the full sample regression on all observations.

It is straightforward to generalize Eq. (7) to the case in which we exclude more than one observation from the regression. To do so, consider the following two regressions:

$$\mathbf{R} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}, \quad (8)$$

$$\mathbf{R} = \mathbf{X}\boldsymbol{\gamma}^{(\Theta)} + \mathbf{E}_\Theta \boldsymbol{\alpha} + \mathbf{u}^{(\Theta)}, \quad (9)$$

where Eq. (8) is the full sample regression of \mathbf{R} on \mathbf{X} , and \mathbf{u} is its residual. Equation (9) is the full sample regression of \mathbf{R} on \mathbf{X} and \mathbf{E}_Θ , where $\mathbf{E}_\Theta = [\mathbf{e}_{\theta_1} | \dots | \mathbf{e}_{\theta_q}]$ and \mathbf{e}_{θ_j} is a $[N \times 1]$ vector with θ_j th element equal to one and all others equal to zero. $\boldsymbol{\gamma}^{(\Theta)}$ and $\boldsymbol{\alpha}$ are the $[(K + 1) \times 1]$ and $[q \times 1]$ parameter vectors of this regression, respectively, and $\mathbf{u}^{(\Theta)}$ is its residual. Because of the dummy variables in \mathbf{E}_Θ , the parameter vector $\boldsymbol{\gamma}^{(\Theta)}$ is equivalent to the parameter vector from the regression of \mathbf{R} on \mathbf{X} excluding the observations in $\Theta = \{\theta_1, \theta_2, \dots, \theta_q\}$.

Pre-multiplying Eq. (9) by $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ gives:

$$\mathbf{P}_X \mathbf{R} = \mathbf{P}_X \mathbf{X} \hat{\boldsymbol{\gamma}}^{(\Theta)} + \mathbf{P}_X \mathbf{E}_\Theta \hat{\boldsymbol{\alpha}} + \mathbf{P}_X \hat{\mathbf{u}}^{(\Theta)}, \quad (10)$$

or, equivalently:

$$\mathbf{X} \hat{\boldsymbol{\gamma}} = \mathbf{X} \hat{\boldsymbol{\gamma}}^{(\Theta)} + \mathbf{P}_X \mathbf{E}_\Theta \hat{\boldsymbol{\alpha}}, \quad (11)$$

where we use the facts that the parameter estimate, $\hat{\boldsymbol{\gamma}}$, is $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}$ and that the residual $\hat{\mathbf{u}}^{(\Theta)}$ is orthogonal to the other exogenous variables. Solving Eq. (11) for the difference between the two parameter estimate vectors, $(\hat{\boldsymbol{\gamma}}^{(\Theta)} - \hat{\boldsymbol{\gamma}})$, gives:

$$\hat{\boldsymbol{\gamma}}^{(\Theta)} - \hat{\boldsymbol{\gamma}} = -(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}_\Theta \hat{\boldsymbol{\alpha}}, \quad (12)$$

where we use the fact that $\mathbf{X}^T \mathbf{P}_X = \mathbf{X}^T$.

Following from the Frisch–Waugh–Lovell theorem, the α estimate can be obtained from the following regression:

$$M_X R = M_X E_\Theta \alpha + r, \tag{13}$$

where r is the residual of this regression. Using the closed-form solution for the estimate, the estimate of the parameter vector α is:

$$\hat{\alpha} = (E_\Theta^T M_X^T M_X E_\Theta)^{-1} (E_\Theta^T M_X^T M_X R) \tag{14}$$

$$= (E_\Theta^T M_X E_\Theta)^{-1} (E_\Theta^T \hat{u}) \tag{15}$$

Substituting Eq. (15) into Eq. (12) gives:

$$\hat{y}^{(\Theta)} - \hat{y} = -(X^T X)^{-1} X^T E_\Theta (E_\Theta^T M_X E_\Theta)^{-1} E_\Theta^T \hat{u}. \tag{16}$$

Equation (16) allows us to calculate the statistical leverage of a set of observations. For example, assume that we want to exclude observations 5, 101, 743, and 3201 from the regression of R on X . Then $q = 4$, $\Theta = \{5, 101, 743, 3201\}$, and $E_\Theta = [e_5 | e_{101} | e_{743} | e_{3201}]$. Plugging E_Θ , the exogenous variable matrix X , and the full sample residual \hat{u} into Eq. (16), we are able to determine the statistical leverage from excluding these four observations.

Using Eq. (16) to identify the set of q observations that most strongly contribute to the parameter estimate vector poses the problem that the number of candidate sets featuring q stocks is $\binom{N}{q}$, easily a very large number. Fortunately, we now show that, as the total number of observations increases relative to the number of excluded observations, the $\hat{y}^{(\Theta)} - \hat{y}$ obtained from jointly excluding the observations in Θ converges to the sum over the $\hat{y}^{(i)} - \hat{y}$ obtained from individually excluding the same observations. Thus, when we exclude a relatively small number of observations, the observations producing the largest joint statistical leverage are also those that have the largest individual statistical leverage.

To prove the above claim, write out $\hat{y}^{(\theta_j)} - \hat{y}$:

$$\hat{y}^{(\theta_j)} - \hat{y} = - \begin{bmatrix} N & \sum_{n=1}^N X_{n,1} & \cdots & \sum_{n=1}^N X_{n,K} \\ \sum_{n=1}^N X_{n,1} & \sum_{n=1}^N X_{n,1}^2 & \cdots & \sum_{n=1}^N X_{n,1} X_{n,K} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{n=1}^N X_{n,K} & \sum_{n=1}^N X_{n,1} X_{n,K} & \cdots & \sum_{n=1}^N X_{n,K}^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \frac{u_{\theta_j}}{1-h_{\theta_j, \theta_j}} \\ \frac{u_{\theta_j} X_{\theta_j,1}}{1-h_{\theta_j, \theta_j}} \\ \vdots \\ \frac{u_{\theta_j} X_{\theta_j,K}}{1-h_{\theta_j, \theta_j}} \end{bmatrix}, \tag{17}$$

where $X_{n,K}$ indicates the n th observation of the k th exogenous variable and $h_{i,j}$ is the row i , column j element of the matrix $P_X = X(X^T X)^{-1} X^T$. Denoting the row i , column j element of the first right-hand side matrix by $\delta_{i,j}$, the $(k + 1)$ th element of $(\hat{\gamma}^{(\theta_j)} - \hat{\gamma})$ is:

$$\hat{\gamma}_{k+1}^{(\theta_j)} - \hat{\gamma}_{k+1} = - \sum_{i=1}^{K+1} \delta_{k+1,i} \frac{u_{\theta_j} X_{\theta_j,i-1}}{1 - h_{\theta_j,\theta_j}}, \tag{18}$$

where $X_{\theta_j,0} \equiv 1$.

Writing out $(\hat{\gamma}^{(\Theta)} - \hat{\gamma})$, we obtain:

$$\begin{aligned} \hat{\gamma}^{(\Theta)} - \hat{\gamma} = & -(X^T X)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{\theta_1,1} & X_{\theta_2,1} & \cdots & X_{\theta_q,1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{\theta_1,K} & X_{\theta_2,K} & \cdots & X_{\theta_q,K} \end{bmatrix} \\ & \times \begin{bmatrix} 1 - h_{\theta_1,\theta_1} & -h_{\theta_1,\theta_2} & \cdots & -h_{\theta_1,\theta_q} \\ -h_{\theta_1,\theta_2} & 1 - h_{\theta_2,\theta_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ -h_{\theta_1,\theta_q} & \cdots & \cdots & 1 - h_{\theta_q,\theta_q} \end{bmatrix}^{-1} \begin{bmatrix} u_{\theta_1} \\ u_{\theta_2} \\ \vdots \\ u_{\theta_q} \end{bmatrix}. \end{aligned} \tag{19}$$

Denoting the row i , column j element of the first right-hand side matrix by $\delta_{i,j}$ and assuming that the off-diagonal elements of M_X are zero, we obtain:

$$\begin{aligned} \hat{\gamma}^{(\Theta)} - \hat{\gamma} = & - \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,K+1} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{K+1,1} & \delta_{K+1,2} & \cdots & \delta_{K+1,K+1} \end{bmatrix} \\ & \times \begin{bmatrix} \frac{u_{\theta_1}}{1-h_{\theta_1,\theta_1}} + \frac{u_{\theta_2}}{1-h_{\theta_2,\theta_2}} + \cdots + \frac{u_{\theta_q}}{1-h_{\theta_q,\theta_q}} \\ \frac{u_{\theta_1} X_{\theta_1,1}}{1-h_{\theta_1,\theta_1}} + \frac{u_{\theta_2} X_{\theta_2,1}}{1-h_{\theta_2,\theta_2}} + \cdots + \frac{u_{\theta_q} X_{\theta_q,1}}{1-h_{\theta_q,\theta_q}} \\ \vdots \\ \frac{u_{\theta_1} X_{\theta_1,K}}{1-h_{\theta_1,\theta_1}} + \frac{u_{\theta_2} X_{\theta_2,K}}{1-h_{\theta_2,\theta_2}} + \cdots + \frac{u_{\theta_q} X_{\theta_q,K}}{1-h_{\theta_q,\theta_q}} \end{bmatrix}. \end{aligned} \tag{20}$$

The $(k + 1)$ th element of $(\hat{\gamma}^{(\Theta)} - \hat{\gamma})$ is then given by:

$$\hat{\gamma}_{k+1}^{(\Theta)} - \hat{\gamma}_{k+1} = - \sum_{i=1}^{K+1} \delta_{k+1,i} \sum_{j=1}^q \frac{u_{\theta_j} X_{\theta_j,i-1}}{1 - h_{\theta_j,\theta_j}} = \sum_{j=1}^q \left(- \sum_{i=1}^{K+1} \delta_{k+1,i} \frac{u_{\theta_j} X_{\theta_j,i-1}}{1 - h_{\theta_j,\theta_j}} \right), \tag{21}$$

where again $X_{\theta_j,0} \equiv 1$.

Comparing Eqs. (18) and (21), we see that $\hat{\gamma}_{k+1}^{(\ominus)} - \hat{\gamma}_{k+1}$ is equal to $\sum_{\ominus} \hat{\gamma}_{k+1}^{(\theta_j)} - \hat{\gamma}_{k+1}$ if the off-diagonal elements of \mathbf{M}_X are zero. In the univariate case (i.e., when \mathbf{X} contains a constant and one regressor), an arbitrary off-diagonal element of \mathbf{M}_X is:

$$-h_{ei,ej} = \frac{-\sum_{n=1}^N X_n^2 + (X_{ei} + X_{ej}) \sum_{n=1}^N X_n - NX_{ei}X_{ej}}{N \sum_{n=1}^N X_n^2 - (\sum_{n=1}^N X_n)^2} \tag{22}$$

$$= \frac{-\frac{1}{N} \sum_{n=1}^N X_n^2 + (X_{ei} + X_{ej}) \frac{1}{N} \sum_{n=1}^N X_n - X_{ei}X_{ej}}{N(\frac{1}{N} \sum_{n=1}^N X_n^2 - (\frac{1}{N} \sum_{n=1}^N X_n)^2)} \tag{23}$$

$$= \frac{1}{N} \frac{-\overline{X^2} + (X_{ei} + X_{ej})\overline{X} - X_{ei}X_{ej}}{\hat{\text{var}}(X)}, \tag{24}$$

where \overline{X} is the sample mean of X , $\overline{X^2}$ is sample mean of X^2 , and $\hat{\text{var}}(X)$ is the sample variance of X . Under standard assumptions, the sample mean and variance converge to constants as the number of observations increases to infinity. Thus, $h_{ei,ej}$ converges to zero as the number of observations increases to infinity, and $\hat{\gamma}_{k+1}^{(\ominus)} - \hat{\gamma}_{k+1}$ to $\sum_{\ominus} \hat{\gamma}_{k+1}^{(\theta_j)} - \hat{\gamma}_{k+1}$. While more tedious, we can prove a similar result when there is more than one regressor in \mathbf{X} .

Appendix 2: Derivation of the run test statistic

We now derive the asymptotic distribution of Chan et al.'s (2003) run test. Assume there are a total of N_t firms in each time period t and a total of T time periods. We denote by $\gamma(i, t)$ the value of a specific firm characteristic for firm i . Define a new random variable $x(i, t)$ equal to one if the firm characteristic value is above the $(1 - p)$ th percentile in period t and else zero:

$$x(i, t) = \begin{cases} 1, & \text{if } \gamma(i, t) > \text{perc}_{1-p}(\gamma(i, t)) \\ 0, & \text{else,} \end{cases} \tag{25}$$

where $\text{perc}_{1-p}(\gamma(i, t))$ is the $(1 - p)$ th percentile of $\gamma(i, t)$ constructed from the N_t firms. Under the null hypothesis that each firm is equally likely to lie above the breakpoint, the random variable $x(i, t)$ follows a simple discrete probability distribution given by:

$$\text{Prob}(x(i, t) = 1) = p \quad \text{and} \quad \text{Prob}(x(i, t) = 0) = 1 - p. \tag{26}$$

Now define another random variable $y(i, t^*, T^*)$:

$$y(i, t^*, T^*) = \prod_{t=t^*}^{T^*} x(i, t), \quad \text{with } t^* \in \{1, \dots, T - l + 1\}, \tag{27}$$

where $y(i, t^*, T^*)$ is equal to one if $x(n, t)$ is equal to one in each time period between t^* and T^* , T^* is defined as $t^* + l - 1$, and l is the length of the test period. Under the null hypothesis of no persistence, the expected value of $y(i, t^*, T^*)$ is given by p^l and its variance is given by $p^l(1 - p^l)$. The cross-sectional mean of $y(i, t^*, T^*)$ is given by:

$$\bar{y}(t^*, T^*) = \frac{1}{N_{t^*, T^*}} \sum_{i=1}^{N_{t^*, T^*}} y(i, t^*, T^*). \quad (28)$$

The cross-sectional mean can be interpreted as the proportion of stocks whose firm characteristic value is consistently above the $(1 - p)$ th percentile during the t^* to T^* -time period.¹⁹

Define a final random variable by deducting p^l from $\bar{y}(t^*, T^*)$ and multiplying by $(N_{t^*, T^*})^{\frac{1}{2}}$:

$$(N_{t^*, T^*})^{\frac{1}{2}} z(t^*, T^*) = (N_{t^*, T^*})^{-\frac{1}{2}} \sum_{i=1}^{N_{t^*, T^*}} (y(i, t^*, T^*) - p^l). \quad (29)$$

Because each summand has an expectation of zero and the summands are only weakly dependent, we can apply an appropriate central limit theorem to the new random variable:

$$\text{plim}_{N_{t^*, T^*} \rightarrow \infty} (N_{t^*, T^*})^{\frac{1}{2}} z(t^*, T^*) \sim N \left(0, \frac{1}{N_{t^*, T^*}} \text{Var} \left(\sum_{i=1}^{N_{t^*, T^*}} (y(i, t^*, T^*) - p^l) \right) \right). \quad (30)$$

We can rewrite the variance of $\sum_{i=1}^{N_{t^*, T^*}} (y(i, t^*, T^*) - p^l)$ as:

$$\sum_{i=1}^{N_{t^*, T^*}} \text{Var}(y(i, t^*, T^*) - p^l) + \sum_{i \neq j} \text{Cov}(y(i, t^*, T^*) - p^l, y(j, t^*, T^*) - p^l). \quad (31)$$

Each variance terms equals $p^l(1 - p^l)$. We can write one of the covariance terms as:

$$\begin{aligned} \text{Cov}(y(i, t^*, T^*) - p^l, y(j, t^*, T^*) - p^l) &= \text{E}[(y(i, t^*, T^*) - p^l)(y(j, t^*, T^*) - p^l)] \\ &= \text{E}[y(i, t^*, T^*) \cdot y(j, t^*, T^*)] - p^{2l}. \end{aligned} \quad (33)$$

¹⁹ If $l = 1$, then $\bar{y}(t^*, T^*)$ is not a random variable, as exactly p percent of its underlying values are equal to unity.

The expected value of $y(i, t^*, T^*) \cdot y(j, t^*, T^*)$ is given by:

$$E[y(i, t^*, T^*) \cdot y(j, t^*, T^*)] \tag{34}$$

$$= E[y(i, t^*, T^*)E[y(j, t^*, T^*)|y(i, t^*, T^*)]] \tag{35}$$

$$= (1 - p)^l \cdot 0 \cdot E[y(j, t^*, T^*)|y(i, t^*, T^*) = 0] + p^l \cdot 1 \cdot E[y(j, t^*, T^*)|y(i, t^*, T^*) = 1] \tag{36}$$

$$= p^l E[y(j, t^*, T^*)|y(i, t^*, T^*) = 1] \tag{37}$$

$$= p^l (\text{Prob} \{y(j, t^*, T^*) = 1|y(i, t^*, T^*) = 1\} \cdot 1 + \text{Prob} \{y(j, t^*, T^*) = 0|y(i, t^*, T^*) = 1\} \cdot 0) \tag{38}$$

$$= p^l \text{Prob} \{y(j, t^*, T^*) = 1|y(i, t^*, T^*) = 1\}. \tag{39}$$

The conditional probability in the last equality can be rewritten as:

$$\text{Prob} \{y(j, t^*, T^*) = 1|y(i, t^*, T^*) = 1\} \tag{40}$$

$$= \text{Prob}\{(x(j, t^*) = 1 \wedge \dots \wedge x(j, T^*) = 1)|(x(i, t^*) = 1 \wedge \dots \wedge x(i, T^*) = 1)\} \tag{41}$$

$$= \text{Prob} \{x(j, t^*) = 1|x(i, t^*) = 1\} \cdot \dots \cdot \text{Prob} \{x(j, T^*) = 1|x(i, T^*) = 1\} \tag{42}$$

$$= \left(\frac{N_{t^*, T^*} p - 1}{N_{t^*, T^*} - 1} \right)^l, \tag{43}$$

where the second equality follows from the facts that the $x(j, t)$ are independent and that the only relevant information in $y(i, t, T) = 1$ for estimating the probability of $x(j, t, T) = 1$ is that $x(i, t, T) = 1$. The third equality follows because there are a total of N_{t^*, T^*} firms of which $N_{t^*, T^*} \times p$ produce a firm characteristic value above the $(1 - p)$ th percentile. If firm i has a firm characteristic value above the $(1 - p)$ th percentile, only $(N_{t^*, T^*} \cdot p - 1)$ of the other $(N_{t^*, T^*} - 1)$ firms can have one too, implying that $\text{Prob} \{x(j, t^*) = 1|x(i, t^*) = 1\} = \left(\frac{N_{t^*, T^*} p - 1}{N_{t^*, T^*} - 1} \right)$.

Because there are a total of N_{t^*, T^*} identical variance terms and a total of $N_{t^*, T^*}(N_{t^*, T^*} - 1)$ identical covariance terms, the variance of the sum in Eq. (31) is given by:

$$N_{t^*, T^*} p^l (1 - p^l) + N_{t^*, T^*} (N_{t^*, T^*} - 1) \left(p^l \left(\frac{N_{t^*, T^*} p - 1}{N_{t^*, T^*} - 1} \right)^l - p^{2l} \right), \tag{44}$$

implying that the variance of the test statistic $z(t^*, T^*)$ is:

$$\text{Var}(z(t^*, T^*)) = \frac{1}{N_{t^*, T^*}} \left(p^l (1 - p^l) + (N_{t^*, T^*} - 1) \left(p^l \left(\frac{N_{t^*, T^*} p - 1}{N_{t^*, T^*} - 1} \right)^l - p^{2l} \right) \right). \tag{45}$$

Computing the cross-sectional mean $\bar{y}(t^*, T^*)$ over several non-overlapping periods, each containing l periods, the time-series average of the cross-sectional mean is normally distributed, with expected value equal to p^l and variance equal to the value given by Eq. (45) divided by the number of non-overlapping periods.

References

- Admati, A., Pfleiderer, P.: A theory of intraday patterns: volume and price variability. *Rev. Financ. Stud.* **1**, 3–40 (1988)
- Amihud, Y.: Illiquidity and stock returns: cross-section and time-series effects. *J. Financ. Mark.* **5**, 31–56 (2002)
- Aretz, K., Bartram, S.M., Pope, P.F.: Macroeconomic risks and characteristic-based factor models. *J. Bank. Financ.* **34**, 1383–1399 (2010)
- Avramov, D., Chordia, T., Jostova, G., Philipov, A.: Credit ratings and the cross-section of stock returns. *J. Financ. Mark.* **12**, 469–499 (2009)
- Avramov, D., Chordia, T., Jostova, G., Philipov, A.: Anomalies and financial distress. Emory Business School Working Paper Series (2011)
- Ball, R., Kothari, S.P., Shanken, J.: Problems in measuring portfolio performance: an application to contrarian investment strategies. *J. Financ. Econ.* **38**, 79–107 (1995)
- Banz, R.W.: The relationship between return and market value of common stocks. *J. Financ. Econ.* **9**, 3–18 (1981)
- Barberis, N., Shleifer, A., Vishny, R.: A model of investor sentiment. *J. Financ. Econ.* **49**, 307–343 (1998)
- Belsley, D.A., Kuh, E., Welsch, R.E.: *Regression Analysis: Identifying Influential Data and Sources of Collinearity*. Wiley, New York (1980)
- Berk, J.B., Green, R.C., Naik, V.: Optimal investment, growth options, and security returns. *J. Financ.* **54**, 1153–1607 (1999)
- Black, F.: Return and the beta. *J. Portf. Manag.* **20**, 8–18 (1993)
- Blume, M.E., Stambaugh, R.F.: Biases in computed returns: an application to the size effect. *J. Financ. Econ.* **12**, 387–404 (1983)
- Boguth, O., Carlson, M., Fisher, A., Simutin, M.: On horizon effects and microstructure bias in average returns and alphas. University of Toronto Working Paper Series (2011)
- Brennan, M.J., Subrahmanyam, A.: Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. *J. Financ. Econ.* **41**, 441–464 (1996)
- Campbell, J., Hilscher, J., Szilagyi, J.: In search of distress risk. *J. Financ.* **63**, 2899–2939 (2008)
- Carhart, M.M.: On persistence in mutual fund performance. *J. Financ.* **52**, 57–82 (1997)
- Chan, K.C., Chen, N.F.: Structural and return characteristics of small and large firms. *J. Financ.* **46**, 1467–1484 (1991)
- Chan, K.C., Chen, N.F., Hsieh, D.A.: An explanatory investigation of the firm size effect. *J. Financ. Econ.* **14**, 451–471 (1985)
- Chan, L.K.C., Jegadeesh, N., Lakonishok, J.: Momentum strategies. *J. Financ.* **51**, 1681–1713 (1996)
- Chan, L.K.C., Karceski, J., Lakonishok, J.: The level and persistence of growth rates. *J. Financ.* **58**, 643–684 (2003)
- Chen, N.F., Roll, R., Ross, S.A.: Economic forces and the stock market. *J. Bus.* **59**, 383–403 (1986)
- Chordia, T., Shivakumar, L.: Momentum, business cycle, and time-varying expected returns. *J. Financ.* **57**, 985–1019 (2002)
- Cochrane, J.H.: *Asset Pricing*. Princeton University Press, New Jersey (2001)
- Cooper, M.J., Gutierrez, R.C., Hameed, A.: Market states and momentum. *J. Financ.* **59**, 1345–1365 (2004)
- Daniel, K., Titman, S.: Evidence on the characteristics of cross sectional variation in stock returns. *J. Financ.* **52**, 1–33 (1997)
- Daniel, K., Grinblatt, M., Titman, S., Wermers, R.: Measuring mutual fund performance with characteristic-based benchmarks. *J. Financ.* **52**, 1035–1058 (1997)
- Daniel, K., Hirshleifer, D., Subrahmanyam, A.: Investor psychology and security market under and over-reactions. *J. Financ.* **53**, 1839–1885 (1998)
- Davidson, R., MacKinnon, J.: *Econometric Theory and Methods*. Oxford University Press, New York (2004)
- Dichev, I.D.: Is the risk of bankruptcy a systematic risk? *J. Financ.* **53**, 1131–1147 (1998)

- Fama, E.F.: *Foundations of Finance*. Basic Books, New York (1976)
- Fama, E.F., French, K.R.: The cross-section of expected stock returns. *J. Financ.* **47**, 427–465 (1992)
- Fama, E.F., French, K.R.: Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **33**, 3–56 (1993)
- Fama, E.F., French, K.R.: Size and book-to-market factors in earnings and returns. *J. Financ.* **50**, 131–155 (1995)
- Fama, E.F., MacBeth, J.D.: Risk, return, and equilibrium: empirical tests. *J. Polit. Econ.* **71**, 607–636 (1973)
- Garlappi, L., Shu, T., Yan, H.: Default risk, shareholder advantage and stock returns. *Rev. Financ. Stud.* **21**, 2743–2778 (2008)
- George, T.J., Hwang, C.-Y.: A resolution of the distress risk and leverage puzzles in the cross section of stock returns. *J. Financ. Econ.* **96**, 56–79 (2010)
- Griffin, J.M., Ji, S., Martin, J.S.: Momentum investing and business cycle risk: evidence from pole to pole. *J. Financ.* **58**, 2515–2547 (2003)
- Hahn, J., Lee, H.: Yield spreads as alternative risk factors for size and book-to-market. *J. Financ. Quant. Stud.* **41**, 247–269 (2006)
- Hong, H., Stein, J.C.: A unified theory of underreaction, momentum trading, and overreaction in asset markets. *J. Financ.* **54**(6), 2143–2184 (1999)
- Jagannathan, R., Wang, Z.: The conditional CAPM and the cross-section of expected returns. *J. Financ.* **51**, 3–53 (1996)
- Jegadeesh, N., Titman, S.: Returns to buying winners and selling losers: implications for stock market efficiency. *J. Financ.* **48**, 65–91 (1993)
- Johnson, T.: Forecast dispersion and the cross section of expected returns. *J. Financ.* **59**, 1957–1978 (2004)
- Kaul, G., Nimalendran, M.: Price reversals: bid-ask errors or market overreaction. *J. Financ. Econ.* **28**, 67–93 (1989)
- Knez, P., Ready, M.: On the robustness of size and book-to-market in cross-sectional regressions. *J. Financ.* **52**, 1355–1382 (1997)
- Kothari, S., Shanken, J., Sloan, R.: Another look at the cross-section of expected returns. *J. Financ.* **50**, 185–224 (1995)
- Kyle, A.: Continuous auctions and insider trading. *Econometrica* **53**, 1315–1335 (1985)
- Lakonishok, J., Shleifer, A., Vishny, R.W.: Contrarian investment, extrapolation and risk. *J. Financ.* **49**, 1541–1578 (1994)
- Lambert, R., Leuz, C., Verrecchia, R.: Accounting information, disclosure, and the cost of capital. *J. Account. Res.* **45**, 385–420 (2007)
- La Porta, R.: Expectations and the cross-section of stock returns. *J. Financ.* **51**, 1715–1742 (1996)
- Lesmond, D., Ogden, J., Trzcinka, C.: A new estimate of transaction costs. *Rev. Financ. Stud.* **12**, 1113–1141 (1999)
- Lesmond, D.A., Schill, M.J., Zhou, C.: The illusory nature of momentum profits. *J. Financ. Econ.* **71**, 349–380 (2004)
- Lewellen, J., Nagel, S.: The conditional CAPM does not explain asset pricing anomalies. *J. Financ. Econ.* **82**, 289–314 (2006)
- Li, X., Brooks, C., Miffre, J.: The value premium and time-varying volatility. *J. Bus. Financ. Account.* **36**, 1252–1272 (2009)
- Malkiel, B.G., Xu, Y.: Idiosyncratic risk and security returns. University of Texas Working Paper Series (2006)
- Merton, R.C.: On the pricing of corporate debt: the risk structure of interest rates. *J. Financ.* **29**, 449–470 (1974)
- Merton, R.C.: A simple model of capital market equilibrium with incomplete information. *J. Financ.* **42**, 483–510 (1987)
- Petersen, M.A.: Estimating standard errors in finance panel data sets: comparing approaches. *Rev. Financ. Stud.* **22**, 435–480 (2009)
- Petkova, R.: Do the Fama-and-French factors proxy for innovations in state variables? *J. Financ.* **61**, 581–612 (2006)
- Queen, M., Roll, R.: Firm mortality: using market indicators to predict survival. *Financ. Anal. J.* **43**, 9–26 (1987)
- Rosenberg, B., Reid, K., Lanstein, R.: Persuasive evidence of market inefficiency. *J. Portf. Manag.* **11**, 9–17 (1985)
- Rousseeuw, P.: Least median of squares regression. *J. Am. Stat. Assoc.* **79**, 871–880 (1984)

- Rousseeuw, P., Van Driessen, K.: Computing LTS regression for large data sets. *Data Min. Knowl. Discov.* **12**, 29–45 (2006)
- Shleifer, A., Vishny, R.W.: The limits of arbitrage. *J. Financ.* **52**, 35–55 (1997)
- Vassalou, M., Xing, Y.: Default risk in equity returns. *J. Financ.* **59**, 831–868 (2004)
- Wurgler, J., Zhuravskaya, E.: Does arbitrage flatten demand curves for stocks? *J. Bus.* **75**, 583–608 (2002)
- Xue, Y., Zhang, M.H.: Fundamental analysis, institutional investment, and limits to arbitrage. *J. Bus. Financ. Account.* **38**, 1156–1183 (2011)
- Zhang, X.F.: Information uncertainty and stock returns. *J. Financ.* **61**, 105–136 (2006)

Kevin Aretz holds an M.Sc. degree in Finance from the University of Maastricht (mid-2004) and a Ph.D. degree in the same subject from the University of Lancaster (end-2008). Prior to joining Alliance Manchester Business School as a senior lecturer (Associate Professor) in September 2011, he worked as a lecturer (Assistant Professor) in Finance at the University of Lancaster. He also worked as an advisor to Old Mutual Asset Management in London until 2011. He was a visiting scholar at Cornell University in the second half of 2014. His research interests lie in the area of asset pricing, corporate finance, and forecasting.

Marc Aretz holds a Diploma degree in Electrical Engineering (mid-2007), an MA degree in Empirical Finance (mid-2009), and a Ph.D. degree in Technical Acoustics (end-2012), all from RWTH Aachen University. In 2012, he joined the Acoustics and Vibration Department at BMW in Munich, where he works in the area of NVH drives.