Rev Quant Finan Acc
https://doi.org/10.1007/s11156-017-0689-y

ORIGINAL RESEARCH

Enhancement of value investing strategies based on financial statement variables: the German evidence

Eero J. Pätäri[1] · Timo H. Leivo[1] · Janne Hulkkonen[2] · J. V. Samuli Honkapuro[1]

Abstract This paper examines the added-value of combining traditional valuation ratios with each other as well as with some financial statement variables in the German stock markets during the 2000–2015 period. The results show that combination pays off and, moreover, that the benefits of combination are greater in Germany than in most other developed stock markets. Particularly, we find strong evidence of the added-value of using Piotroski's *F*-score as a supplementary selection criterion for value stocks as well as for low-accrual stocks. Our results show further that the *F*-score also boosts the efficacy of other valuation ratios besides the book-to-price ratio. In addition, the inclusion of *F*-score besides a relative value measure tends to increase the average market equity of portfolio firms. The decomposition of the full-sample-period performance into separate bull- and bear-period performance shows clearly that the better performance of *F*-score-boosted portfolios is mostly attributable to their outperformance during bearish periods, even though on average, they also generate higher bull-period returns than the comparable value portfolios formed without *F*-score. The use of *F*-score as a supplementary criterion also increases the proportion of stocks that earn above-market-average returns during the subsequent holding period. For the first time in the financial literature, we also document a strong relationship between high *F*-score stocks and momentum stocks.

✉ Eero J. Pätäri
eero.patari@lut.fi

Timo H. Leivo
timo.leivo@lut.fi

Janne Hulkkonen
janne.hulkkonen@rainmaker.fi

J. V. Samuli Honkapuro
samuli.honkapuro@lut.fi

[1] School of Business and Management, Lappeenranta University of Technology, P.O. Box 20, 53851 Lappeenranta, Finland

[2] Rainmaker Numbers, Vattuniemenkatu 21, 00210 Helsinki, Finland

Published online: 04 December 2017

Keywords Value premium · Valuation multiples · Value investing · Value anomaly · Accrual anomaly

JEL Classification G11 · G14

1 Introduction

Considerable international evidence of value anomalies, which refer to the tendency of value stocks to outperform the stock market average most of the time, has been documented (e.g., Chan and Lakonishok 2004; Brown et al. 2008a; Fama and French 2006, 2012; Cakici et al. 2013; Pätäri et al. 2017b). This evidence has shown not only that the value anomalies in stock markets are a worldwide phenomenon but also that the relative efficacy of different valuation criteria varies across stock markets and the sample periods examined.[1] In addition, some scholars have started to examine whether the performance of value stock portfolios can be further enhanced by adding other criteria besides relative value to the portfolio selection. These studies can be classified into two categories, the first of which either combines value indicators with momentum indicators (e.g., Pätäri et al. 2012, 2017a; Fisher et al. 2016) or double-sorts the stocks based on their value and momentum indicators (e.g., see Bird and Casavecchia 2007; Leivo 2012). In the second category, test designs attempt to pick the best-performing value stocks of the future from a larger set of value stocks based on some other firm characteristics, such as profitability, accruals or financial strength. The results of the studies of the latter type have been promising as, on average, value stocks with good profitability, a high earnings quality and/or a sound financial condition have performed better than stocks that are just the cheapest in terms of valuation ratios.[2] This study contributes to the latter category by combining these three quality dimensions with a value dimension by using German stock market data. Motivated by the evidence for the added-value of combining individual valuation ratios into composite value criteria,[3] we also test whether the benefits of creating combinations within the value dimension are comparable to those achievable by combining the value and quality dimensions. For this purpose, we form some simple 2-combination value portfolios. We also examine the added-value of the inclusion of Piotroski's (2000) *F*-score as the supplementary criterion in cases of forming such composite value portfolios.

In general, the benefits of combining selection criteria are related to the inter-relationship between the criteria being compared. The weaker the relationship, the better the potential for the added-value of combining. For this reason, the added-value of combining individual valuation ratios is generally deemed to be limited, as they all are reasoned to represent the same (i.e., value) dimension (e.g., see Fama and French 2011). However, recent evidence has shown that the performance of value portfolios formed on different valuation ratios does vary (e.g., see Barbee et al. 2008; Gray and Vogel 2012; Pätäri et al. 2017b), and therefore, we include several valuation ratios in our analysis. Among these, *operating cash flow-to-price* (CFO/P) is particularly interesting, as Desai et al. (2004) conclude that it can capture the mispricing attributes of both the value and the accrual

---

[1] E.g., see Pätäri and Leivo (2017) for a comprehensive literature review on value anomalies and the reasons behind them.

[2] E.g., see Novy-Marx (2013) for evidence on the benefits of combining value and profitability criteria, Bartov and Kim (2004) and Simlai (2016) for the benefits of combining value and earnings quality (determined on the basis of accruals in these two studies) criteria, and Piotroski (2000), Piotroski and So (2012) and Novy-Marx (2014) for the added-value of combining value and financial strength indicators.

[3] E.g., see Pätäri et al. (2016) and Leshem et al. (2016).

anomalies, even though these two anomalies capture distinct sources of mispricing.[4] Another motivation for the inclusion of CFO/P is that despite these striking results by Desai et al. (2004), CFO/P is a relatively infrequently examined valuation ratio in the related literature, although the results of the few studies in which it has been employed for value portfolio selection have been favorable for this specific value measure (see Dissanaike and Lim 2010 for U.K. evidence, and Kim et al. 2012 for U.S. evidence). To find out whether the accrual anomaly found by Sloan (1996) is subsumed by the value anomaly when using CFO/P as a proxy for relative value in the German stock market as well, we also include accruals as one of portfolio-formation criteria in our analysis.

Motivated by the results of Novy-Marx (2013), we also examine the added-value of combining value and profitability dimensions for portfolio selection purposes. Instead of the gross profitability employed by Novy-Marx, we use return-on-equity (ROE) as a measure of profitability, because ROE is more closely related to the B/P ratios that we use as proxies for relative value in combining value and profitability dimensions[5] (Because the tendency of low (high) B/P firms to be more (less) profitable (in terms of ROE) than high (low) B/P firms is well-documented (e.g., see Damodaran 2012; Penman and Reggiani 2013), there is a more solid theoretical foundation for such value-profitability combinations due to the inverse relationship between B/P and ROE[6]).

As the third quality indicator, we employ Piotroski's *F*-score that, based on nine binary signals designed to measure three different dimensions (i.e., profitability, change in financial leverage/liquidity, and change in operational efficiency) of firms' financial strength, forms a composite quality score within a range of 0–9. Recent evidence has shown that coupled with B/P rankings, Piotroski's *F*-score is among the most efficient commonly-used quality criteria in building value-quality equity portfolios (e.g., see Piotroski and So 2012; Novy-Marx 2014). It is also widely used in the investment industry. However, to the best of our knowledge, the performance of equity investing strategies based on Piotroski's approach has not been examined in the German stock market in academic journal articles prior to our paper.

The potential added-value of using quality indicators beside value indicators stems from the fact that many value stocks remain cheap (in terms of relative value) for an extended period of time. In such cases, the investor may buy them much too early, which leads to a poor portfolio performance. However, when the quality criterion/criteria is/are also taken into account in stock selection decisions, the probability of poorly timed purchases

[4] The branch of accounting literature that discusses whether the accrual anomaly is distinct and incremental to other pricing anomalies is abundant (e.g., Collins and Hribar (2000) conclude that the accrual anomaly is distinct from the post-earnings announcement drift (PEAD) anomaly, whereas Barth and Hutton (2004) show that the predictive ability of accruals for future returns is not subsumed by the predictive ability of analysts' forecast revisions. By contrast, Dechow et al. (2008) conclude that the accrual anomaly subsumes the investing and the external financing anomalies, whereas Ball et al. (2016) state that cash-based operating profitability subsumes accruals in predicting the cross section of average returns). The same also holds for the literature on the reasons and explanations for the accrual anomaly (e.g., see Richardson et al. 2010, and Dechow et al. 2011, for extensive reviews on such studies).

[5] According to a decomposition of B/P, it can be determined on the basis of the dividend growth model of Gordon and Shapiro (1956) and the relationship between the sustainable growth rate (g), dividend payout ratio (DIV/EPS) and ROE as follows: Assuming that $P = DIV/(r - g)$, and $\mathrm{ROE} = EPS/B$, and $g = ROE\,(1 - DIV/EPS)$, it follows that $B/P = (r - g)/(ROE - g)$. *Ceteris paribus*, higher ROE justifies lower B/P, and vice versa.

[6] E.g., Haugen and Baker (1996), Novy-Marx (2013) and Hou et al. (2017) show that firms with higher ROE earn higher subsequent returns than firms with lower ROE.

decreases, which boosts the overall performance of the portfolio.[7] Thus, the quality indicators help investors detect which of the value stocks are actually underpriced and which of them are cheap for a reason. Because the market responds slowly to accounting information (e.g., see Ou and Penman 1989), quality indicators based on financial statements are useful for assessing whether a cheap-looking stock should be bought immediately or later, or whether it should not be bought at all.

The German stock market provides an interesting setting for this type of study, as the value anomaly has been documented to be exceptionally weak there (e.g., see Capaul et al. 1993; Fama and French 1998; Artmann et al. 2012b). Therefore, the added-value of using financial statement information for the performance enhancement of value portfolios may be greater in such circumstances. Our results show that this is actually the case. We also contribute to the existing literature by coupling for the first time, to our best knowledge, Piotroski's *F*-score with several other valuation measures besides *book-to-price* (B/P). The performance of the *F*-score-boosted portfolios is also compared with that of the equal-sized quantile portfolios formed on the same valuation measures.[8] In addition, we test the added-value of combining *F*-score with low-accrual stocks.

The structure of the paper is as follows: Sect. 2 describes the data and methodology, while Sect. 3 introduces the empirical results. Section 4 concludes.

2 Data and methodology

2.1 Sample selection and methodology

The portfolios are composed of non-financial German stocks quoted in the Frankfurt Stock Exchange (FSE). The sample period extends from May 2000 to April 2015. To be included in the portfolios, the firms had to have financial statement variables available for two previous fiscal years preceding each portfolio-formation point. To avoid survivorship bias, the sample also includes the stocks of the companies that were delisted during the observation period. If an issuer has had two or more stock series listed, only the one with the higher trading volume is included in the sample.[9] The firm-years with fiscal year ends in months other than December are excluded from the sample.[10] Adjustments for dividends, splits and capitalization issues are made appropriately. In line with the seminal *F*-

[7] The addition of the second-layer quality criteria beside a value criterion can also boost the so-called hit rates that indicate the proportion of stocks whose returns have been higher than the return of the benchmark portfolio. For example, by using a value-weighted market index as the benchmark, Piotroski (2000) reports positive market-adjusted 1-year holding-period returns for only 43.7% of the top-quintile B/P stocks, whereas the corresponding hit rate for the portfolio that consists of the highest *F*-score stocks (i.e., scores of either 8 or 9) in the highest B/P quintile is 50.0%.

[8] In spite of numerous robustness tests, Piotroski (2000) does not compare the performance of the portfolio formed from the highest *F*-score stocks picked from the top-quintile B/P stocks to the quantile portfolio that would have consisted of an equal number of the highest B/P stocks.

[9] One of the peculiarities of the German stock market is the dual class system, where many firms have two stock series, common and non-voting stocks that are often misinterpreted as preferred stocks and excluded from the samples in academic studies. However, this should not be done, as even for some major German firms, the non-voting stocks may compose the integral part of their market equity. Moreover, in some cases their trading volume can be higher than that of the common stocks of the same company (see Brückner 2013 for details).

[10] We set this limitation to ensure that all firm-year observations are equally fresh when forming the value portfolios.

score study of Piotroski (2000), we do not set any percentile market-cap breakpoint for the inclusion condition.[11] Instead, we control the size effect by including the size factor (SMB) in the 4-factor model, which is used as a basis for determining abnormal returns for each portfolio. SMB also reveals whether the portfolios are, on average, tilted towards small- or large-cap stocks. Moreover, at each portfolio-formation checkpoint, the stocks whose prices are below €1.00 and which have a market-cap below €5 million are excluded from the universe of investable stocks for the subsequent 1-year holding period.[12] The stock return data is from the Thomson Reuters Datastream,[13] whereas financial statement data is from Worldscope. One-month EURIBOR (downloaded from Datastream) is used as a proxy for the risk-free rate of return.

To be included in the sample of investable stocks at each portfolio-formation point, the firms must have all the information available for the calculation of all 13 selection criteria being examined. Although this prerequisite reduces the number of otherwise usable firm-year observations, it allows the best possible comparison of the results based on different single selection criteria and/or combination criteria, consistent with Dhatt et al. (2004) and Pätäri et al. (2017a). In line with the existing literature (e.g., see Fama and French 1992, 2008, Artmann et al. 2012a; Gharghori et al. 2013), we also exclude firm-year observations for which the book value of equity is negative.[14] After all exclusion criteria, the final sample size ranges from 230 companies in the year 2000 to 465 in 2008, comprising 5713 firm-year observations, with complete data for each of the 13 selection criteria over the 2000–2015 sample period.

Accounting data is from the latest available financial statements published prior to the moments of rebalancing, whereas the market values of equity used in the denominators of the price multiples are updated to match those prevailing at the end of April in year t for all firms.[15] The sample stocks are first ranked based on valuation multiples, accruals, market leverage or combination criteria calculated on every rebalancing date, that is, the first trading day of May, with a 1-year frequency.[16] The stocks are then divided into decile portfolios based on these ranks. In the first stage, we use 12 different selection criteria for

[11] Piotroski (2000) documents the benefits of F-score to be greatest among small- and mid-cap firms in the U.S. stock markets (Novy-Marx 2014 also ends up with parallel results with a longer dataset ranging from mid-1963 to the end of 2012). As we, unlike Piotroski, also compare the performance of F-score-boosted portfolios with that of the value portfolios of the same size (in terms of the number of stock series in the portfolios), our research design allows us to also examine the impact of the inclusion of the F-score on the robustness of the results to size effect. Another reason for not establishing a percentile market-cap breakpoint is the dual stock classes of many German firms described in footnote 9 (In some cases, the Datastream market equity values for German dual class firms are based only on the market equity of common shares and, therefore, erroneously omit the market equity part of non-voting stocks).

[12] We set the market-cap threshold besides the penny stock filter because not all German penny stocks are simultaneously small-caps (for examples see Brückner 2013).

[13] Being aware of the problems in Datastream equity return data (e.g., see Ince and Porter 2006), we checked and corrected the total return time series for each equity according to the guidelines suggested by Brückner (2013). As a result of such actions, our sample is comprehensive and representative, given that the coverage of the German stocks in Datastream has increased dramatically since 1990 (According to Brückner 2013, a complete total return time series is available for 97.92% of German stocks after 1990).

[14] See also Brown et al. (2008b) and Ang (2015) for discussions on the characteristics of negative book equity stocks.

[15] We follow this practice because we aimed to use the latest available information in portfolio formation without any look-ahead bias consistent with Lakonishok et al. (1994), Desai et al. (2004), Asness et al. (2013), and Fisher et al. (2016), among others.

[16] Consistent with previous literature (e.g., Lakonishok et al. 1994; Dhatt et al. 2004; Gharghori et al. 2013; Fisher et al. 2016), we form portfolios 4 months after the end of the fiscal year to avoid look-ahead bias.

ranking the stocks into decile portfolios. Three of these are traditional price multiples (i.e., B/P, *earnings yield* (E/P) and *sales-to-price* (S/P)). The three other single criteria being examined are *operating cash flow-to-price* (CFO/P), employed by Desai et al. (2004), Dissanaike and Lim (2010), and Kim et al. (2012), *market leverage*, employed by Fama and French (1992), Loughran and Wellman (2011) and Artmann et al. (2012a), and *accruals-to-assets*, used in Piotroski (2000), Dopuch et al. (2010), and Lee and Lee (2015). To the best of our knowledge, the inter-relationship between the value and accrual anomalies has not been examined previously in the German stock market on the basis of portfolio sorts.

The remaining combination criteria are those that have been proven or can be supposed to be efficient on the basis of previous financial literature. The first of these is the combination of E/P and B/P, in which the average of the B/P and E/P rankings at each portfolio-formation point determines the rank score for each firm (The same methodology is applied in forming all the combination strategies that will be introduced next). The second is the corresponding combination of B/P and return on equity (ROE). The relationship between B/P and ROE has often been discussed in the valuation literature (e.g., see Wilcox 1984; Penman 1991; Leibowitz 1999; Wilcox and Philips 2005). In addition, Piotroski (2000) finds that high B/P firms with the highest profitability earn higher returns than comparable firms with the lowest profitability. Therefore, it is interesting to examine whether the double criterion formed on these two variables is more efficient than B/P alone in detecting undervalued stocks.

Motivated by the findings of Bartov and Kim (2004) and Simlai (2016), we also include the combination of accruals-to-assets and B/P as one portfolio-formation criterion (The authors documented the increase in both value portfolio return and value premium when the value portfolio was selected from high B/P stocks with a low ratio of accruals-to-assets). The fourth combination in which B/P is the other selection criterion is the combination of B/P and S/P, motivated by the results of Bird and Whitaker (2003).[17] Based on the previous literature (e.g., see Dhatt et al. 2004), the S/P criterion is also combined with each of the earnings multiples (i.e., E/P and CFO/P). Altogether, six combination criteria are used for ranking the stocks into decile portfolios.

In the second part of the empirical analysis, we examine the same portfolio-formation criteria as in the comparison of decile portfolios, but we include *F*-score as a supplementary criterion. We first form value tertile portfolios based on each criteria. In the second stage, the top-tertile stocks with the two highest *F*-scores (i.e., either 8 or 9) are picked for the ultimate value portfolios. To better abstract the added-value of *F*-score, we also form comparable quantile portfolios in which the number of constituent stocks is set equal to that of the highest *F*-score stocks in the corresponding tertile portfolio for each holding period.[18] Similar to Hyde (2016), we also form a portfolio of high *F*-score stocks without any connection to their relative value, although Piotroski's (2000) original purpose

[17] Based on the large sample of stocks from seven major European developed countries (Germany included), the authors find that high B/P and S/P portfolios consisted of different type of stocks; whereas high B/P quintile stocks were mostly below average in terms of market cap, high S/P firms were clearly bigger than high B/P firms. In addition, the high S/P quintile consists of stocks for which the average B/P was close to the B/P middle-quintile average.

[18] For example, when the number of high *F*-score stocks (determined on the basis of 1999 financial statements) included in the E/P top-tertile in spring 2000 is 23, the ultimate E/P value quantile portfolio being compared against the combined E/P and *F*-score portfolio consists of the same number of the highest E/P stocks during the subsequent holding period.

for using F-score was to select from the larger set of value stocks those that are financially strong in spite of their low relative value.

The stocks included in each portfolio are equal-weighted at each portfolio-formation point. The weight changes of the stocks stemming from their return differences within 1-year holding periods are taken into account in the calculation of monthly quantile portfolio returns.[19] The calculation principles for the valuation ratios, accruals and market leverage are described in Appendix 1, whereas Appendix 2 introduces the binary components of F-score, which are determined analogously to Piotroski (2000).

2.2 Sample characteristics

Descriptive statistics of the sample stocks (Table 1) indicates that the levels of relative value are clearly dependent on the overall stock market conditions and vary remarkably over time. Based on the cross-sectional medians of B/P, CFO/P, S/P, and market leverage as end of April, the German stocks were at their cheapest in terms of their relative value in spring 2003,[20] just before the long bullish period that preceded the financial crisis. However, based on the comparable medians of E/Ps, the lowest relative end-of-April valuation took place in 2009, just after the first relief of the financial crisis. By contrast, the valuation criteria disagree more on the point of the highest relative valuation: Based on B/P and CFO/P medians, it was in spring 2007 just before the financial crisis, whereas based on S/P and market leverage, the average relative valuation levels were at their highest 1 year later. E/P deviates from other valuation criteria also in this respect, as the lowest E/P median appeared in spring 2004.

The decrease in the median market equity value of the sample periods during the first years of the sample period is explained by several reasons: First, the general stock market decline from the end of April 2000 till the end of April 2003 cut off, on average, 54.9% of the market equities of German stocks (Actually, the maximum drawdown during the burst of the dot-com bubble from the 7th of March, 2000 to the 12th of March, 2003 was − 67.3%). Second, many IPOs were implemented during the late 1990s as well as during the early 2000s, and the issuers were mostly smaller-cap firms or at least ended up to small-caps during the burst of the bubble. For example, by June 2003, the NEMAX All Share, which was the composite index of the Neuer Markt,[21] where most of IPOs around the millennium change had taken place (Brückner 2013), had fallen to only 5% of its all-time-

[19] This methodology is followed instead of using value-weighted returns because from the viewpoint of practical implementation of the investing strategies, our approach is more realistic and easier to implement (Based on the previous literature, according to which both value and accrual anomalies are stronger on an equal-weighted basis than on a value-weighted basis (e.g., see Loughran and Wellman 2011 and Taylor and Wong 2012, respectively), it is also more likely that a portfolio manager would rather equal-weight than value-weight the constituent stocks when deciding on portfolio allocation).

[20] The reader should note that the descriptive statistics is based on only 15 cross-sectional snapshots, and therefore, it only indicates the levels as end of April each year.

[21] The Neuer Markt was an exchange regulated market at the FSE. It was opened in March, 1997 to raise capital to young ICT companies. In many respects, it was comparable to the NASDAQ in New York, the AIM in London, or the Nouveau Marché in Paris. Although the Neuer Markt was very successful in the beginning (e.g., see Vitols 2001), many irregularities and the poor performance of its constituent companies over the 3-year period from March 2000 to March 2003 severely damaged its reputation. From the peak of 8559 points recorded in March 2000, the NEMAX All Share gradually fell to its all-time-low of 358 points in March 2003. As a consequence, the Neuer Markt was closed in June, 2003. However, as the Neuer Markt was not a separate segment of the FSE in legal terms, the constituent companies stayed with their formal listing at the FSE (Stehle and Schmidt 2015).

Table 1 Descriptive statistics

Year	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	ME	# of firms
2000	0.504	10.67	3.98	2.137	− 0.043	1.799	204.16	230
2001	0.494	6.70	4.59	1.826	− 0.025	1.541	204.74	252
2002	0.621	7.42	2.43	1.936	− 0.060	1.798	125.43	350
2003	0.933	15.57	2.82	2.939	− 0.074	2.638	83.52	320
2004	0.648	11.06	2.18	1.925	− 0.070	1.796	55.26	409
2005	0.561	11.07	4.15	1.744	− 0.062	1.526	54.04	399
2006	0.524	7.73	4.02	1.360	− 0.038	1.305	75.56	397
2007	0.457	6.09	4.50	1.291	− 0.020	1.242	88.61	406
2008	0.475	6.60	4.41	1.243	− 0.031	1.216	83.06	442
2009	0.806	11.65	5.91	2.304	− 0.044	2.105	45.13	456
2010	0.629	10.57	3.06	1.596	− 0.063	1.636	61.50	437
2011	0.557	8.53	4.98	1.371	− 0.033	1.356	85.05	418
2012	0.695	8.52	5.73	1.671	− 0.033	1.660	73.83	420
2013	0.652	8.73	4.58	1.632	− 0.043	1.527	86.95	400
2014	0.508	7.84	3.79	1.359	− 0.044	1.242	103.24	377
Average	0.604	9.25	4.08	1.756	− 0.046	1.626	95.34	380

For each of the six single portfolio-formation criteria examined, the table presents annual medians, as well as their full-sample period averages (The medians of CFO/P and E/P are in percentages). The second-last column provides the corresponding statistics on market equity values (ME) of the sample firms (in million euros). The right-most column shows the number of the sample firms in each 1-year sub-period, whereas the left-most column indicates the time points (as end of April each year) for which the statistics are calculated (see Appendices 1 and 2 for calculation principles for each characteristic). *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage

high recorded in March 2000 (Stehle and Schmidt 2015). The turbulence caused by the financial crisis triggered the similar-type of stock market crash than the burst of the technology bubble a few years earlier, although the later crash was not so deep in Germany than was the earlier crash (The maximum drawdown during the crash caused by the financial crisis from the 9th of July, 2007 to the 6th of March, 2009 was − 50.5%). In addition, as the stock market index was at the higher level before the financial crisis than it was before the burst of the technology bubble, the financial crisis does not explain alone, why the median market equity of the sample firms as end of April 2009 is even at lower level than it had previously been as end of April within the sample period. The other reason for this is that the new entrants have unsurprisingly been smaller firms than those already included in the sample. Despite the fact that during the last 6 years of the sample period, the market-cap median has risen remarkably, it was still only approximately half of what it was in the beginning of the sample period.

Based on the pooled annual cross-sections of firm-year variables employed in portfolio formation, Table 2 shows the pairwise variable correlations over the 15-year holding period. The strongest correlation is between the S/P and market leverage ratios, whereas the weakest (in terms of absolute values) is reported between B/P and ACCR. The strongest negative correlation is between the E/P and market leverage ratios. Generally, the strongest correlations with the other variables are documented for E/P, but the corresponding correlation coefficients vary a lot, being outstandingly negative in some cases,

Table 2 The cross-sectional Pearson correlation coefficients for the portfolio-formation variables

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*
CFO/P	0.288*					
E/P	0.258*	0.231*				
S/P	0.007	0.051*	− 0.306*			
ACCR	− 0.002	− 0.041*	0.140*	− 0.020		
MLEV	− 0.025**	0.088*	− 0.501*	0.713*	− 0.020	
F-score	− 0.004	0.237*	0.215*	− 0.042*	0.020	− 0.085*

Based on the pooled annual cross-sections of firm-year variables employed in portfolio formation, table shows the pairwise variable correlations over the 15-year sample period from April 2000 to April 2014. The significance of the correlations is shown by asterisks. * (**) indicates a significant correlation at the 1% (10%) level. *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage

while positive in some other cases. Somewhat surprisingly, the correlation between E/Ps and S/Ps is highly negative, whereas between E/P and ACCR, it is positive, in contrast with the correlations between ACCR and the value indicators other than E/P[22] (The expected cross-sectional relation between ACCR and the other included variables is reverse to that for the other pairs of variables because according to the accruals anomaly, low-ACCR stocks should outperform high-ACCR stocks, whereas for the other variables, the stocks at the high end of variable distributions should outperform the corresponding low-end stocks, given that the variables have anomalous return relations).

With respect to the cross-sectional correlations of the other portfolio-formation variables with the *F*-score, the most significant of these is documented between CFO/P and F-score, whereas the most insignificant is reported for B/P and *F*-score. The latter finding suggests that the added-value of boosting value portfolio performance with the *F*-score may be the highest when a value portfolio is formed on B/P (This is actually the case, as shown by our results). Generally, due to the high number of firm-year observations the correlations are highly significant in most cases in spite of low absolute values of the related coefficients.[23]

Based on monthly return time-series over the 15-year holding period, Table 3 presents the pairwise return correlations for the value decile portfolios formed on the single selection criteria. The correlation triangle below the diagonal of the correlation matrix shows the pairwise Pearson correlations between the decile portfolios formed on high-B/P, -CFO/P, -E/P, -S/P, and -MLEV stocks, and low-accrual stocks. As expected, all the correlation coefficients are highly significant, ranging from 0.655 between the top-decile E/P portfolio and the bottom-decile accruals portfolio to 0.879 between the top-decile portfolios formed on B/P and market leverage. On average, the top-decile E/P correlates the least with the other five decile portfolios being compared, whereas the top-decile CFO/P portfolio correlates the most with the other five. However, the correlation differences

[22] To find out whether the positive correlation between E/Ps and ACCRs is caused by negative E/P stocks, we divided the firm-year observations into two groups based on the signs of E/Ps and calculated the corresponding correlations separately for each group. The resulting coefficients remained significantly positive for both groups, thereby indicating that this somewhat unexpectable cross-sectional relation between E/P and ACCR is not driven by negative E/P stocks.

[23] With 5711 degrees of freedom, the ranges for the insignificant correlation coefficients are very narrow: At the 10% (1%) significance level, the range is from − 0.022 (− 0.034) to 0.022 (0.034), approximately.

Table 3 The Pearson correlation coefficients for the value decile portfolios formed on the single selection criteria

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*
B/P	**0.437**	*0.840*	*0.745*	*0.840*	*0.791*	*0.779*
CFO/P	0.817	**0.677**	*0.809*	*0.828*	*0.834*	*0.797*
E/P	0.708	0.774	**0.755**	*0.767*	*0.712*	*0.758*
S/P	0.754	0.813	0.717	**0.549**	*0.786*	*0.905*
ACCR	0.825	0.790	0.655	0.714	**0.522**	*0.700*
MLEV	0.879	0.814	0.737	0.796	0.734	**0.416**

Based on monthly return time-series over the 15-year holding period from May 2000 to April 2015, the table presents the pairwise return correlations for the value decile portfolios formed on the single selection criteria. The correlation triangle below the diagonal of the correlation matrix shows the pairwise Pearson correlations between the decile portfolios formed on high-B/P, -CFO/P, -E/P, -S/P, and -MLEV stocks, and low-accrual stocks, whereas the triangle above the diagonal shows the correlations (denoted in italics) between the corresponding *F*-score-boosted quantile portfolios. The diagonal indicates the correlations (in bold) between the *F*-score-boosted quantile portfolios and the comparable plain value portfolios that consists of the same number of stock series as included in the corresponding *F*-score-boosted quantile portfolios. *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage

between the decile portfolios are very small, as the average correlations calculated for each of the decile portfolios range from 0.718 to 0.802. Based on the pairwise correlations, the return-generation pattern of the low-accrual decile portfolio is very similar to the patterns of the top-decile value portfolios. In this respect, our results from the German stock market are in line with the U.S. results of Beaver (2002), Desai et al. (2004) and Simlai (2016), who all conclude that the low-accrual firms are mostly value firms in disguise.

The correlation triangle above the diagonal shows the correlations between the corresponding *F*-score-boosted quantile portfolios. These correlations are very similar to those reported for the decile portfolios formed on single selection criteria, although marginally higher on average. By contrast, the correlations between the *F*-score-boosted quantile portfolios and the comparable plain value portfolios that consist of the same number of stock series as included in the corresponding *F*-score-boosted quantile portfolios (reported on the diagonal of Table 3) are weaker (averaging 0.559) and vary more, ranging from 0.416 (in the case of market leverage) to 0.755 (in the case of E/P). Although these correlation coefficients are highly significant even at their weakest, they show the potential of the *F*-score inclusion for changing the content of the portfolios more than could be done by combining some of the single selection criteria into a composite value indicator.

2.3 Test procedures for performance comparisons

The performance of quantile portfolios is evaluated based on four performance metrics: the raw return, the standard Sharpe ratio (Sharpe 1966), the skewness- and kurtosis-adjusted Sharpe ratio (henceforth SKASR), developed by Pätäri (2011), and the 4-factor alpha (Carhart 1997). We prefer the SKASR to the standard Sharpe ratio in our analysis because the return distributions of the quantile portfolios being examined are not consistent with the normality assumption related to the use of standard deviation as a risk proxy, as in the original Sharpe ratio.[24] The inclusion of higher moments of return distributions in the performance evaluation of equity portfolios is also motivated by several recent studies (e.g., Zakamouline and Koekebakker 2009; Alles and Murray 2010; Homm and Pigorsch 2012; Lee and Su 2012; Feunou et al. 2013; Theodossiou and Savva 2016). In addition, the comparison between results based on the standard Sharpe ratio and those based on the SKASR show that, in some cases, the shapes of the return distributions of the quantile portfolios being examined deviate from normality to the extent that their performance rank orders change when allowing for the skewness and kurtosis dimensions in performance evaluation. To avoid validity problems stemming from the negative excess returns in the context of the SKASR comparisons, we use a similar refinement in the denominator of the ratio as suggested by Israelsen (2005) for the negative Sharpe ratios:

$$SKASR = \frac{r_i - r_f}{SKAD_i^{(ER/|ER|)}} \tag{1}$$

where r_i is the average monthly return of portfolio i, r_f is the average monthly risk-free rate of return, $SKAD_i$ is the skewness- and kurtosis-adjusted standard deviation of the monthly excess returns of portfolio i and ER is the average excess return of portfolio i.

SKAD captures the third and fourth moments of the return distributions being analyzed. Based on fourth-order Cornish–Fisher (1937) expansion, the adjusted Z-value (i.e., Z_{CF}) that corresponds to the Z-value of normal distribution is first calculated as follows:

$$Z_{CF} = Z_C + \frac{1}{6}\left(Z_C^2 - 1\right)S + \frac{1}{24}\left(Z_C^3 - 3Z_C\right)K - \frac{1}{36}\left(2Z_C^3 - 5Z_C\right)S^2 \tag{2}$$

where Z_c is the critical value of the probability based on standard normal distribution, and S refers to Fisher's skewness and K to excess kurtosis of the return distribution. Next, we calculate SKAD by multiplying the standard deviation by the ratio Z_{CF}/Z_c. Consistent with Favre and Galéano (2002) and Pätäri and Tolvanen (2009), we set Z_c to -1.96 to correspond to a 95% probability level when determining this ratio.

To evaluate whether the potential abnormal returns are explained by four commonly used explanatory factors, we also calculate 4-factor alphas for each quantile portfolio based on the following regression equation:

$$r_{it} - r_{ft} = \alpha_i + b_i(r_{mt} - r_{ft}) + s_i SMB_t + h_i HML_t + m_i WML_t + \varepsilon_{it} \tag{3}$$

where r_{it} is the return of a portfolio, r_{ft} is the risk-free rate of return, α_i is the 4-factor alpha (the abnormal return over and above what might be expected based on the 4-factor model employed), r_{mt} is the stock market return, SMB_t is the return of the size factor (i.e., the return difference between small- and large-cap portfolios), HML_t is the return of the B/P factor (i.e., the return difference between high and low B/P portfolios), WML_t is the return of the momentum factor (i.e., the return difference between winner and loser stock portfolios), b_i, s_i, h_i and m_i are factor sensitivities to the stock market, SMB, HML and WML factors, respectively, and ε_i is the residual term.

Because of the weighting system employed, we use equal-weighted returns calculated from the comprehensive sample of German stocks provided by Stehle et al. (2015) as a

[24] Although we use the SKASR as our primary measure of total risk-adjusted performance, we report the standard Sharpe ratios in tables and also discuss the discrepancies between these two performance metrics if they produce inconsistent rankings for the portfolios being compared (The reader should note that if the return distributions being compared were strictly normal, both types of Sharpe ratios, as well as the corresponding risk metrics, would be exactly equal in such cases).

proxy for the market return. The other factors are also calculated on the basis of equal-weighted returns,[25] otherwise following the methodology employed by Fama and French (1993) for the construction of the SMB and HML factors and that of Fama and French (2012) for the construction the WML factor.

The statistical significance of the differences between comparable pairs of total-risk-adjusted returns is given by the *p* values of the Ledoit–Wolf test,[26] which is based on the circular block bootstrap method. We also test the significance of 4-factor alphas based on their *t*-statistics. In addition, we test the statistical significance of the differences between the quantile portfolio alphas using the appropriate alpha spread test (Pätäri et al. 2010). Throughout the study, we use Newey–West (1987) standard errors in the statistical tests to avoid problems related to autocorrelation and heteroscedasticity.

3 Results

3.1 The results for the decile portfolios

Table 4 shows the results for the top-decile portfolios. Of the six single selection criteria examined, the best one is CFO/P, which generates both the highest return (17.94% p.a.) and the highest SKASR (0.861) as well as the highest 4-factor alpha (4.58% p.a.). Based on the total-risk adjusted statistics, CFO/P significantly outperforms the stock market portfolio (at the 0.01% level), whereas based on the 4-factor alpha, its outperformance is only mildly significant (at the 10% level).[27] The worst single selection criterion in terms of both raw and total-risk adjusted returns is the accruals-to-assets, which generates the lowest return (5.59% p.a.) with the highest risk. However, based on the 4-factor alphas, the worst within the same peer group is S/P, for which the alpha is negative, although insignificantly (− 3.14% p.a.).

Although S/P does not work well as a stand-alone criterion for the purpose of value portfolio selection in the German stock market, it is effective when combined with earnings multiples. Among all 12 top-decile portfolios, the best two are the combinations of S/P and E/P and of S/P and CFO/P. The latter is the best in terms of both raw returns (19.67%

[25] After careful consideration of various factor sets, we find that this type of factor produces the highest adjusted R-squareds as well as the lowest alphas for the portfolios being examined (see Brückner et al. 2015 for an analytic comparison of alternative German factor sets). The comprehensive data set provided by Stehle et al. (2015) covers all three segments of the FSE and is consistent with the German stock universe available from Datastream. Of the two alternative quantile return time-series for each quantile-division criterion calculated with and without the corporate tax credit, we chose the latter since the incremental return stemming from the stock owners' compensation of the corporate tax that the firm has paid for dividend payouts is not taken into account in total return calculations of equities in Datastream. In addition, such an imputation credit was available only for the German shareholders for dividends paid by the German companies until the system was discontinued in October 2000 (e.g., see Stehle and Schmidt 2015 for details). The quantile time-series employed in the calculation of factor returns are downloadable at: https://www.wiwi.hu-berlin.de/de/professuren/bwl/bb/data/fama-french-factors-germany/fama-french-factors-for-germany.

[26] The Ledoit–Wolf test also takes account of skewness and kurtosis of return distributions being compared, as does also the SKASR. Because of the complexity of the test procedure as well as space limitations, we do not describe the Ledoit–Wolf test in more detail here, but we recommend the original article to the interested reader (Ledoit and Wolf 2008). The corresponding programming code is freely available at http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9.

[27] For all the 12 decile portfolios, the significance of the outperformance over the market portfolio is higher based on the SKASR difference than based on the 4-factor alphas.

Table 4 Performance of value decile portfolios

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Market
RR (%)	10.62	17.94	14.75	7.73	5.59	11.49	15.85	12.35	16.22	12.10	19.67	17.88	4.81
σ (%)	21.73	18.35	15.57	19.08	24.09	20.16	15.12	18.34	15.64	23.52	18.70	15.63	17.45
SKAD (%)	21.35	18.04	18.19	18.93	24.32	19.51	17.45	18.91	18.38	20.64	17.56	17.46	20.21
SR	0.383	0.842	0.793	0.287	0.141	0.454	0.887	0.545	0.881	0.415	0.917	0.985	0.150
SKASR	0.391	0.861	0.682	0.291	0.140	0.472	0.773	0.531	0.754	0.475	0.981	0.886	0.131
Sign. (%)	(13.5)	(0.0)	(0.5)	(41.0)	(95.8)	(7.7)	(0.0)	(2.6)	(0.1)	(6.0)	(0.0)	(0.0)	
4-F-alpha (%)	0.29	4.58	4.17	− 3.14	− 0.81	− 1.53	5.33	− 0.19	4.95	0.33	6.59	8.52	
Sign. (%)	(91.81)	(9.83)	(19.42)	(31.92)	(81.21)	(55.98)	(0.85)	(92.86)	(6.90)	(90.42)	(0.53)	(0.16)	
Beta	1.175	1.087	0.849	1.027	1.228	1.144	0.862	1.073	0.903	1.274	1.061	0.825	
Sign. (%)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	
SMB	0.487	0.140	− 0.076	0.193	0.578	0.179	− 0.085	0.155	− 0.071	0.563	0.134	− 0.144	− 4.03
Sign. (%)	(0.2)	(10.7)	(52.7)	(20.9)	(0.0)	(11.7)	(47.7)	(21.5)	(45.0)	(0.4)	(20.3)	(17.5)	
HML	0.393	0.209	0.157	0.172	0.294	0.276	0.184	0.273	0.148	0.520	0.191	0.104	9.13
Sign. (%)	(0.2)	(6.9)	(1.3)	(0.7)	(0.6)	(0.2)	(0.4)	(0.3)	(2.6)	(0.5)	(1.9)	(0.7)	
WML	0.151	0.259	0.161	0.237	0.061	0.261	0.133	0.230	0.180	0.166	0.249	0.102	20.81
Sign. (%)	(6.9)	(0.0)	(0.1)	(0.1)	(40.8)	(0.0)	(0.7)	(0.1)	(0.0)	(12.0)	(0.0)	(0.9)	
Adj. R^2	69.8	69.5	61.8	58.3	73.0	64.3	71.1	69.5	69.0	69.9	64.0	64.7	

The table presents the annualized geometric average returns (RR), volatilities (σ), SKADs (all in percentages), the Sharpe ratios (SR), the SKASRs, 4-factor alphas and betas, slopes for SMB, HML, and WML factors, as well the adjusted R-squareds for the 10 top-decile portfolios formed on the valuation criteria named on the column header line, as well as for the bottom-decile accruals-to-assets *(ACCR)* portfolio and for the decile portfolio formed on the combination of high B/P and low-accrual stocks, over the May 2000–April 2015 period. The right-most column shows the corresponding statistics for the market portfolio, if appropriate, and annualized average premiums (in percentages) for the SMB, HML and WML factors. The significance levels are in percentages in parentheses [The sign. levels for the SKASR differences indicate the significances for the outperformance over the stock market portfolio, whereas the t-statistics of the alphas indicate the significances for abnormal return over the 4-factor asset pricing model introduced in Eq. (3)]. The average number of constituent stocks in the decile portfolios is 38, ranging from 23 during the May 2000–April 2001 period to 46 during the May 2009–April 2010 period. *MLEV* refers to market leverage (i.e., total assets/market value of equity). All the statistics are calculated on the basis of monthly returns p.a.) and SKASR, whereas the former generates the highest and most significant alpha (8.52% p.a. that is significant at the 0.2% level). In contrast, the combination of S/P and B/P does not perform as well, since its 4-factor alpha is slightly though insignificantly negative (− 0.19% p.a., being the only negative alpha among the six combination criteria examined). Its raw return (12.35% p.a.) is also much lower than that of the combination of S/P and E/P and that of S/P and CFO/P. It is also slightly riskier than the latter two combinations in terms of both total and systematic risk. Altogether, the comparison of all 12 decile portfolios shows that the best combinations can add value to the investor, although none of them significantly outperform CFO/P, which is the best of the six single criteria in terms of both raw and risk-adjusted returns.

Based on adjusted R-squareds, the 4-factor model explains reasonably well the excess returns of the top-decile portfolios. The lowest adjusted R-squared (58.35%) is found for the portfolio formed on S/P, whereas the highest (72.97%) is generated by the portfolio formed on accruals-to-assets. The highest market exposure (market beta of 1.274) is reported for the combination portfolio of B/P and accruals-to-assets, and the lowest exposure (0.825) is documented for the combination of S/P and E/P. The size factor (SMB) is significant in only 3 out of 12 cases (i.e., for B/P, accruals-to-assets, and their combination), being positive in all these 3 significant cases. It is the most significant for the top-decile accruals-to-assets portfolio, implying that this particular portfolio has the strongest tilt towards small-cap stocks.[28] Given that the size premium (SMB) is negative during the sample period,[29] the small-cap bias largely explains the poor performance of the bottom-decile accruals-to-assets portfolio even though the return variability of the same portfolio is also significantly affected by the HML factor. The most significant exposure to the latter factor is expectedly documented for B/P, which is used as the basis for calculating HML.

By contrast, the most significant exposure to the momentum (WML) factor is documented for the top-decile CFO/P portfolio, with a *t*-statistic as high as 5.64. Somewhat surprisingly, the momentum is a significant explanatory factor for 9 out of 12 top-decile portfolios,[30] although it is not included in the portfolio-formation criteria. This finding underlines the importance of the inclusion of WML in the factor-based performance evaluation in cases of value portfolios as well.

3.2 The results for the F-score-boosted portfolios and the comparable quantile portfolios

Table 5 shows that the combination of S/P and CFO/P also performs well in cases where the number of stocks in value portfolios is determined by the number of stocks with the two highest *F*-scores (i.e., scores of 8 and 9) in each value tertile. Due to a positively skewed and exceptionally leptokurtic return distribution, this combination generates the highest SKASR among all the examined portfolios, including the one formed on the basis of plain *F*-score criteria as well as those in which *F*-score is used as a supplementary criterion (see Table 6). This finding also highlights the importance of allowing for higher moments in total risk-adjusted performance measurement, as the combination S/P and CFO/P generates

[28] This finding is consistent with the Australian evidence of Clinch et al. (2012) and Taylor and Wong (2012) as well as with the U.S. evidence of Lev and Nissim (2006), who all find that the extreme accrual firms are mostly small-caps.

[29] The reverse size effect in the German stock market has also been documented in earlier periods (e.g., see Stehle 1997; Brückner et al. 2012).

[30] The three exceptions are the same for which the SMB factor is significant.

Table 5 Performance of value quantile portfolios and plain F-score portfolio

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Plain *F*-score	Market
RR (%)	8.87	15.50	15.08	8.99	1.48	18.64	13.00	10.32	15.84	4.55	22.04	18.07	18.19	4.81
σ (%)	27.55	19.92	15.77	23.21	28.65	30.58	16.18	24.70	17.61	30.78	23.66	16.10	15.06	17.45
SKAD (%)	24.76	19.35	18.21	21.46	28.23	24.33	17.88	22.97	19.57	21.70	14.09	17.25	17.40	20.21
SR	0.240	0.657	0.803	0.290	− 0.002	0.528	0.659	0.325	0.763	0.077	0.824	0.969	1.040	0.150
SKASR	0.268	0.679	0.699	0.314	− 0.002	0.666	0.598	0.350	0.689	0.110	1.388	0.908	0.907	0.131
Sign. (%)	(56.2)	(0.0)	(1.5)	(46.8)	(49.5)	(4.9)	(1.9)	(34.8)	(0.6)	(91.9)	(0.0)	(0.0)	(0.0)	
4-F-alpha (%)	2.46	2.56	2.29	− 1.56	− 2.28	5.28	2.39	− 3.43	4.28	− 2.04	8.45	8.02	4.81	
Sign. (%)	(65.67)	(40.30)	(54.32)	(75.59)	(69.86)	(24.89)	(44.15)	(35.48)	(20.11)	(76.18)	(1.97)	(0.36)	(2.92)	
Beta	1.081	1.099	0.839	0.989	1.230	1.268	0.860	1.215	0.949	1.265	1.053	0.802	0.932	
Sign. (%)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	
SMB	0.640	0.209	− 0.117	0.236	0.760	0.502	− 0.071	0.234	0.000	0.800	0.144	− 0.085	− 0.137	− 4.03
Sign. (%)	(0.0)	(1.0)	(38.9)	(11.4)	(1.1)	(2.4)	(58.4)	(19.9)	(99.7)	(0.5)	(50.0)	(52.6)	(15.2)	
HML	0.420	0.204	0.196	0.289	0.293	0.455	0.240	0.368	0.170	0.544	0.144	0.141	0.118	9.13
Sign. (%)	(0.1)	(9.7)	(1.0)	(4.8)	(4.9)	(9.7)	(0.1)	(0.8)	(3.5)	(0.3)	(10.6)	(3.1)	(7.9)	
WML	0.072	0.280	0.240	0.220	0.032	0.301	0.138	0.300	0.204	0.065	0.319	0.134	0.256	20.81
Sign. (%)	(68.0)	(0.0)	(0.0)	(1.3)	(80.9)	(3.5)	(1.9)	(0.4)	(0.0)	(73.5)	(0.1)	(1.1)	(0.0)	
Adj. R^2	41.2	59.4	52.5	35.9	55.8	34.3	60.9	46.9	58.7	46.8	35.7	53.3	72.5	

Table 5 continued

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Plain *F*-score	Market
n	17	28	28	16	20	16	29	17	23	20	25	26	52	

The table presents the annualized geometric average returns (RR), volatilities (σ), SKADs (all in percentages), the Sharpe ratios (SR), the SKASRs, 4-factor alphas and betas, slopes for SMB, HML, and WML factors, as well the adjusted R-squareds for the quantile portfolios formed on the 13 portfolio-formation criteria (named on the column header line) over the May 2000–April 2015 period. The right-most column shows the corresponding statistics for the market portfolio, if appropriate, and annualized average premiums (in percentages) for the SMB, HML and WML factors. The significance levels are in percentages in parentheses [The sign. levels for the SKASR differences indicate the significances for the outperformance over the stock market portfolio, whereas the t-statistics of the alphas indicate the significances for abnormal return over the 4-factor asset pricing model introduced in Eq. (3)]. The number of constituent stocks in each quantile portfolio equals the number of the stocks with two highest F-scores (i.e., either 8 or 9) in the corresponding top-tertile (bottom-tertile) portfolio formed on each value or composite measure (or *ACCR*) and varies over 1-year holding periods. However, the plain F-score portfolio is exceptional in that its constituent stocks (i.e., those with two highest F-scores) are picked from the whole universe of investable stocks without any tertile restrictions. The bottom row shows the average number of constituent stocks calculated over the full-length 15-year holding period for each portfolio-formation criterion. *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage

Table 6 Performance of F-score-boosted value portfolios

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Market
RR (%)	23.92	22.26	19.00	23.54	20.91	21.17	20.01	23.98	21.14	23.11	22.71	21.34	4.81
σ (%)	19.72	17.64	15.49	19.42	20.52	18.23	15.98	20.09	17.33	19.78	16.94	15.88	17.45
SKAD (%)	17.67	18.87	19.01	18.86	18.14	16.31	18.57	18.24	18.94	14.87	18.21	17.96	20.21
SR	1.080	1.114	1.061	1.078	0.896	1.020	1.092	1.063	1.071	1.038	1.185	1.181	0.150
SKASR	1.211	1.048	0.871	1.115	1.017	1.147	0.946	1.177	0.986	1.386	1.110	1.051	0.131
Sign. (%)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	
4-F-alpha (%)	9.00	9.27	6.77	9.11	5.79	7.88	6.40	8.68	8.45	8.16	7.95	7.66	
Sign. (%)	(0.69)	(1.19)	(1.76)	(0.06)	(18.89)	(0.13)	(1.59)	(0.67)	(3.31)	(3.37)	(0.03)	(0.06)	
Beta	0.981	0.912	0.884	1.021	1.083	0.898	0.916	1.052	0.884	1.057	0.990	0.930	
Sign. (%)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	
SMB	0.063	− 0.168	− 0.162	− 0.058	− 0.131	− 0.146	− 0.080	0.110	− 0.047	0.175	− 0.154	− 0.167	− 4.03
Sign. (%)	(70.2)	(21.5)	(13.6)	(64.0)	(59.8)	(18.6)	(42.0)	(43.4)	(68.7)	(32.0)	(11.0)	(9.0)	
HML	0.123	0.100	0.134	0.093	0.153	0.130	0.135	0.133	0.152	0.143	0.130	0.135	9.13
Sign. (%)	(5.8)	(5.6)	(7.5)	(15.2)	(9.9)	(1.0)	(4.1)	(2.1)	(2.7)	(4.8)	(0.2)	(2.0)	
WML	0.342	0.234	0.199	0.306	0.315	0.252	0.266	0.353	0.230	0.346	0.281	0.241	20.81
Sign. (%)	(0.1)	(2.1)	(0.0)	(0.0)	(0.5)	(0.0)	(0.0)	(0.0)	(0.5)	(0.1)	(0.0)	(0.0)	
Adj. R^2	42.7	51.5	65.0	50.7	51.4	45.0	60.5	48.3	49.0	51.2	63.9	66.1	

Table 6 continued

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Market
n	17	28	28	16	20	16	29	17	23	20	25	26	

The table presents the annualized geometric average returns (RR), volatilities (σ), SKADs (all in percentages), the Sharpe ratios (SR), the SKASRs, 4-factor alphas and betas, slopes for SMB, HML, and WML factors, as well the adjusted R-squareds for the F-score-boosted portfolios formed on the 12 portfolio-formation criteria (named on the column header line) over the May 2000–April 2015 period. The right-most column shows the corresponding statistics for the market portfolio, if appropriate, and annualized average premiums (in percentages) for the SMB, HML and WML factors. The significance levels are in percentages in parentheses [The sign. levels for the SKASR differences indicate the significances for the outperformance over the stock market portfolio, whereas the t-statistics of the alphas indicate the significances for abnormal returns over the 4-factor asset pricing model introduced in Eq. (3)]. The number of constituent stocks in each quantile portfolio equals the number of stocks with the two highest F-scores (i.e., either 8 or 9) in the corresponding top-tertile (bottom-tertile) portfolio formed on each value or composite measure (or *ACCR*) and varies over 1-year holding periods. The bottom row shows the average number calculated over the full-length 15-year holding period for each portfolio-formation criterion). *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage lower standard Sharpe ratio than any of the 12 *F*-score-boosted portfolios included in Table 6. Among the latter group of portfolios, the *F*-score-boosted combination of B/P and accruals-to-assets generates SKASR that is almost as high as for the previously mentioned combination, and in addition, a raw return of 23.11% p.a. These results highlight the incremental value of *F*-score: although the plain B/P criterion and the accruals-to-assets criterion as well as their combination all perform relatively poorly as stand-alone criteria,[31] their performance is dramatically enhanced by adding *F*-score as a supplementary criterion. The most dramatic improvement is in the accruals-to-assets portfolio: among the 37 portfolios examined, it is the only portfolio for which the average return over the 15-year sample period is lower than the risk-free return (see Table 5). Without *F*-score, its return is the lowest and its SKAD the highest, whereas coupling it with *F*-Score increases the average annual return by 19.43 percentage points while simultaneously reducing the total risk dramatically (from 28.23 to 18.14% in terms of SKAD). The same tendency is also seen in the case of the B/P criterion, for which the average annual return increases from 8.87 to 23.92% when combined with *F*-Score. This shows that, at least in the German stock market and for the sample period employed, the added-value of using *F*-score within the high B/P quantile is not explained by the lower number of constituent stocks in the double-criteria portfolio than in the comparable B/P portfolio. Given the strong *F*-score boost documented for both B/P and accruals portfolios, it is not surprising that the combination portfolio formed on these two criteria also benefits greatly from the addition of *F*-score as a supplementary criterion.

Altogether, the efficacy of almost all 12 portfolio-formation criteria can be improved with the addition of *F*-score. The incremental return from its inclusion is positive in all cases, ranging from 0.67% p.a. for the combination of S/P and CFO/P to 19.43% p.a. for the accruals-to-assets criterion. The positive boost of *F*-score also holds in terms of SKASR for all criteria except for the combination of S/P and CFO/P. As Table 5 demonstrates, *F*-score also performs relatively well as a stand-alone criterion, as only 2 out of the 12 quantile portfolios formed without the *F*-score criterion can beat it in terms of raw returns, SKASRs, and the *t*-statistics of the 4-factor alpha. The highest overall return (23.98% p.a.) among all 37 examined portfolios is generated by the *F*-score-boosted combination of B/P and S/P. However, it should be noted that the *F*-score-boosted portfolios are not equal in respect to their breadth dimension, as the number of high *F*-score stocks at each portfolio-formation point varies within the tertiles from which such stocks are picked. Over the 15-year sample period, the average number of stock series in these portfolios varies from 16 (for the S/P and market leverage portfolios) to 29 (for the combination of B/P and E/P), with an average of 22.

In most cases, the positive added-value of using *F*-score as a supplementary selection criteria also holds for 4-factor alphas, except for the combinations of S/P and E/P, and S/P and CFO/P. Based on this performance metric, the biggest improvement is documented for the combination of B/P and S/P, which without *F*-score generates a negative alpha of − 3.43% p.a., but coupled with *F*-score achieves a highly significant positive alpha of 8.68% p.a. (the corresponding alpha spread is also significant at the 5% level). The highest overall 4-factor alpha (9.27% p.a.) is generated by the *F*-score-boosted CFO/P portfolio (see Table 6). Somewhat surprisingly, this portfolio is followed by the *F*-score-boosted S/P and B/P portfolios (for which the alphas are 9.11 and 9.00% p.a., respectively). In contrast,

[31] In this sense, our results for German stocks contradict the U.S. evidence of Bartov and Kim (2004) and Simlai (2016), according to whom the combination of high B/P and low-accrual boosted the annual return in comparison with the returns from the corresponding single selection criteria.

the corresponding alphas for the comparable S/P- and B/P-based stand-alone portfolios are − 1.56 and 2.46% p.a., which are parallel to the corresponding decile alphas (and all insignificant). The enhancement of alphas due to the inclusion of *F*-score is also evident in the cases of E/P and CFO/P (for which the "incremental" alphas are 4.48 and 6.71% p.a., respectively). These findings highlight the added-value of coupling *F*-score with the traditional valuation ratios in the German stock market. However, the same does not hold for the combinations of S/P and CFO/P or S/P and E/P, which generate the two highest SKASRs without *F*-scores. For them, the incremental alphas are slightly negative, indicating that the inclusion of *F*-score cannot add value, in terms of alpha percentages, in these two cases.[32] In contrast, for the combination of B/P and accruals-to-assets, the coupling with *F*-score is highly beneficial, increasing the 4-factor alpha from − 2.04 to 8.16% p.a. and removing the highly significant small-cap exposure. The significance of the small-cap tilts also disappears when the portfolios formed on either B/P or accruals-to-assets are boosted by *F*-score.

The range of the adjusted R-squareds is wider for quantile portfolios whose size is determined by the number of high *F*-score stocks in each value tertile than for the decile portfolios. Given higher variability in the average number of stocks in the former type of portfolios, this finding is not surprising. Within the former peer group, the lowest adjusted R-squared (34.31%) is reported for the market leverage portfolio, whereas the highest (66.11%) is documented for the *F*-score-boosted combination of S/P and E/P. The average adjusted R-squareds are also lower in this case (i.e., 51.12% compared to 66.35% in the case of decile portfolios). However, the total risks of *F*-score-boosted portfolios, on average, are slightly lower than those of decile portfolios, which are, in turn, slightly less risky (in terms of total risk) than the quantile portfolios formed without *F*-score. Also within the latter peer group, the lowest exposure to the market factor (0.802) is documented for the combination of S/P and E/P, whereas the highest (1.268) is generated by the market leverage criterion. The wide range of the market slopes indicates that the differences in market exposures are outstanding even between the plain value portfolios. Based on the *t*-statistics of the regression slopes, the B/P quantile portfolio has the strongest exposure to the SMB factor. The size factor also significantly explains the return variability of the corresponding portfolios formed on CFO/P, accruals-to-assets, market leverage, and the combination of B/P and accruals-to-assets, whereas for the comparable *F*-score-boosted portfolios (Table 6), SMB exposures are never significant (at the 5% level). In light of these results, it seems that *F*-score can, if not totally remove, at least alleviate the small-cap bias that is inherent to portfolios formed on certain single selection criteria (based on the overall results for this particular sample data, the B/P and accruals-to-assets portfolios seem to be the most prone to this kind of bias).

The HML factor is significant (at the 5% level) for 9 out of the 12 quantile portfolios formed without *F*-score (Table 5), whereas for the comparable *F*-score-boosted portfolios, the number of significant cases decreases to 7 (Table 6). In line with the decile results, momentum is a significant factor for 9 out of the 12 quantile portfolios formed without *F*-score. However, for the 12 *F*-score-boosted quantile portfolios, it is significant (at the 5% level) in all cases, implying that the observed *F*-score boost is clearly related to the momentum spread return. This inference is reinforced by the 4-factor regression results for the plain *F*-score portfolio, in which the momentum factor is highly significant while both

[32] However, the *t*-statistics of the alphas are higher for the corresponding *F*-score-boosted combinations than for the comparable quantile portfolios in all 12 cases, including the combinations of S/P and CFO/P, and S/P and E/P.

the SMB and HML factors are insignificant [the adjusted R-squared is also high (i.e., 72.53%)]. As additional evidence of the interrelationship between the high *F*-score and momentum stocks, we calculate the correlation between the monthly returns of the plain high *F*-score portfolio and an equally weighted momentum quantile portfolio of a similar size, and find it to be extremely significant (the Pearson correlation coefficient is 0.728, with a *t*-statistic as high as 14.18).[33] To the best of our knowledge, such strong evidence of the connection of the *F*-score and price momentum has not been documented in previous literature.[34]

3.3 Anatomy of outperformance of value portfolios

The overall results show that the value investor can benefit from combining valuation ratios[35] as well as from using financial statement information beyond valuation ratios in the German stock market. Based on the decile results, the two best value strategies in terms of both SKASR and 4-factor alphas are the combinations of S/P and CFO/P and of S/P and E/P.[36] The former generates the highest raw return (19.67% p.a.) as well as the highest SKASR (0.981), whereas the latter generates the highest alpha (8.52% p.a.). Unlike the former, the latter has negative SMB exposure, which makes it more easily implementable due to the better liquidity of larger-cap stocks. While both S/P and CFO/P decile portfolios are somewhat inclined to small-caps, the corresponding E/P portfolio is not, and consequently, the combination of S/P and E/P is actually slightly more tilted towards large-cap stocks than the combination of S/P and CFO/P is tilted towards small-caps. The same phenomenon is also observable in the comparison of the corresponding quantile portfolios, the sizes of which are determined on the basis of the number of high *F*-score stocks in the

[33] The return for the equally weighted (EW) momentum quantile portfolio was approximated on the basis of the corresponding momentum decile returns provided by Stehle et al. (2015) as follows: If the proportion of the high *F*-score stocks in the full sample at the portfolio-reformation point was 22%, for example, the corresponding momentum return was determined by weighting the average returns of the three highest EW momentum decile returns so that the top-2 decile returns were given the weight 10/22, whereas the weight for the third highest decile return was 2/22. The same approximation technique, which assumes the returns within each decile to be even, was followed if the proportion was between 10 and 20% (In that case, only the top-2 EW momentum decile returns are relevant). If the proportion of the high *F*-score stocks was below 10%, the top decile return was used per se. On average, the high *F*-score portfolio included 13.54% of the investable stocks, ranging from 7.71% in 2001 to 24.06% in 2004.

[34] Although in his seminal work Piotroski (2000) discusses the possibility of the correlation of *F*-score and momentum, he does not test such a correlation directly. Instead, he runs a 6-factor cross-sectional regression, in which momentum and *F*-score are used as two explanatory factors. Based on the regression results, he concludes that the prediction power of *F*-score on future returns is robust to the momentum effect. However, the results of Piotroski (2000) are not comparable to ours in this respect, as his robustness test is performed for the sample of high B/P stocks, whereas the high *F*-score stocks in our plain *F*-score portfolio are picked from the whole universe of investable stocks at each portfolio-formation point. Moreover, Piotroski's momentum indicator is different from ours (while he uses prior 6-month market-adjusted buy-and-hold return without any lag, we use 12-month returns by skipping the return for the most recent month, as is the convention in the current momentum literature (see, e.g., Fama and French 2008, 2012; Novy-Marx 2012; Asness et al. 2013; Israel and Moskowitz 2013; Cakici et al 2014; Barroso and Santa-Clara 2015).

[35] In this sense, our results deviate from those of Bird and Casavecchia (2007), who, for the pan-European sample data of 7 major developed national markets including Germany over the 1989–2004 sample period, find no improvement in performance from combining valuation ratios.

[36] In this respect, our results are parallel to those of Dhatt et al. (2004), who also document excellent performance for the combination of S/P and E/P for U.S. stocks over the 1980–1998 period. However, in contrast to our results, S/P also worked well as a stand-alone criterion in the U.S. stock markets during Dhatt et al.'s sample period.

corresponding value tertile portfolios (However, the SMB slopes are less significant in this case). In contrast, when *F*-score is added as a supplementary criterion, the small-cap exposure of the combination of S/P and CFO/P disappears, turning into a nearly significant large-cap exposure, implying that *F*-score changes the content of that particular combination quite dramatically. The same tendency of *F*-score inclusion to reduce the small-cap exposure also holds for many other strategies; without *F*-score, the SMB factor is significant and positive for 5 out of the 12 quantile portfolios, whereas it is insignificant for all the *F*-score-boosted portfolios. Coupled with the fact that the returns of *F*-score-boosted portfolios are always higher than those of the corresponding quantile portfolios formed without *F*-scores, this finding provides strong evidence for the added-value of the use of *F*-score. Moreover, based on the average adjusted R-squareds, the *F*-score-boosted portfolios appear to be somewhat better diversified than the comparable quantile portfolios formed without *F*-scores (53.78 vs. 48.46%, respectively).

However, there are some potential pitfalls in *F*-score-boosted portfolios: for example, the number of firms with high *F*-scores varies outstandingly over time. Generally, the worse the state of the economy, the lower the proportion of high *F*-score stocks, and therefore the less diversified the *F*-score-boosted portfolios. We examine the impact of this characteristic of *F*-score-boosted portfolios on their performance by dividing the sample period into bull and bear market periods according to the turning points of the German stock market. We follow Edwards et al. (2003) in using a 20% cumulative return (loss) from the previous trough (peak) to the subsequent peak (trough) in the demarcation of bullish (bearish) periods (see Table 7 for details). Based on the DAX100 index, we identify 4 separate bullish and 4 bearish periods within the 15-year sample period. Panel A in Table 7 shows that the lower diversification of *F*-score-boosted portfolios during the bearish periods is advantageous rather than disadvantageous, as the average monthly losses in such conditions are much smaller for *F*-score-boosted portfolios than for the comparable quantile portfolios formed without *F*-score. While the average monthly loss for the 12 latter type of portfolios (as well as for the 12 value decile portfolios) is − 2.37% during the bearish periods, it is only − 1.11% for the *F*-score-boosted portfolios during the same periods. Moreover, the range of losses is far narrower for *F*-Score-boosted portfolios (from − 0.85 to − 1.44% p.m.) than it is for the comparable quantile portfolios and the decile portfolios (i.e., from − 1.21 to − 4.57% p.m. for the former, and from − 1.41 to − 4.27% p.m. for the latter). The finding is intuitive in the sense that during hard times, investors may favour firms that are financially strong because such firms are less risky due to their stronger financial buffers.[37]

The success of the *F*-score-boosted portfolios is explained not only by their superior performance during bearish periods but also because they generate higher returns during bullish conditions, on average (see Panel B in Table 6). The average monthly return for the 12 *F*-score-boosted portfolios during the bullish periods is 2.88%, whereas it is 2.47% for the comparable quantile portfolios and 2.56% for the decile portfolios. Consistent with the corresponding results in bearish conditions, the range of returns in bullish conditions is much narrower for *F*-score-boosted portfolios (from 2.72 to 3.03% p.m.) than it is for the

[37] In addition, we also split the full-length sample period into bullish months and bearish months on the basis of the signs of stock market average returns in each month, similar to Fuller and Goldstein (2011) [This demarcation criterion identified 105 (75) months with positive (negative) returns]. The results of this analysis are in line with the bullish- and bearish-period results and show even clearer evidence of the downside protection provided by the *F*-score-boosted portfolios (see Panel C in Table 6).

Table 7 Decomposition of the performance of portfolios

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Plain *F*-score	Market
Panel A: Bear cycle														
Decile portfolio	− 3.00	− 1.92	− 1.42	− 2.85	− 4.27	− 2.83	− 1.41	− 2.46	− 1.57	− 3.17	− 1.87	− 1.60		
Quantile portfolio	− 2.84	− 2.32	− 1.21	− 1.95	− 4.57	− 2.71	− 1.44	− 2.93	− 1.48	− 4.04	− 1.60	− 1.30	− 1.42	− 3.57
F-score-boosted portfolio	− 1.01	− 1.30	− 1.44	− 1.07	− 1.05	− 1.05	− 1.19	− 0.85	− 1.15	− 0.95	− 1.12	− 1.10		
Panel B: Bull cycle														
Decile portfolio	2.54	2.83	2.28	2.15	2.55	2.56	2.39	2.48	2.50	2.78	2.99	2.69		
Quantile portfolio	2.27	2.76	2.22	1.89	2.20	3.26	2.10	2.47	2.42	2.33	3.11	2.57	2.64	2.14
F-score-boosted portfolio	3.03	3.00	2.73	3.02	2.75	2.78	2.72	2.97	2.82	2.93	2.96	2.82		
Panel C: Bear months														
Decile portfolio	− 1.73	− 2.95	− 2.51	− 2.22	− 2.75	− 3.78	− 1.79	− 1.78	− 3.13	− 2.45	− 1.66	− 1.89		
Quantile portfolio	− 1.51	− 3.04	− 2.32	− 2.53	− 2.19	− 4.09	− 2.03	− 2.02	− 3.95	− 3.17	− 1.55	− 1.52	− 1.28	− 3.52
F-score-boosted portfolio	− 1.25	− 0.47	− 0.99	− 0.81	− 0.73	− 1.52	− 0.94	− 0.84	− 0.85	− 0.53	− 0.90	− 0.92		
Panel D: Bull months														
Decile portfolio	3.27	3.64	2.92	4.04	3.61	3.59	3.45	3.49	3.98	3.50	3.61	4.01		
Quantile portfolio	3.14	3.48	2.95	3.97	4.11	3.25	3.26	3.62	3.58	3.77	3.55	4.02	3.37	3.28

Table 7 continued

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Plain *F*-score	Market
F-score-boosted portfolio	3.44	3.46	3.80	3.51	3.32	3.88	3.33	3.40	3.64	3.51	3.46	3.65		

For each of the 37 portfolios examined, Panel A reports the monthly geometric average returns for the observed bear market periods, whereas Panel B does the same for the bull market periods. The demarcation between the bear and bull markets is based on the turning points of the DAX100 index (Using a 20% cumulative return (loss) from the previous trough (peak) to the subsequent peak (trough) in the demarcation, 4 separate bearish and 4 bullish periods are identified. The bearish periods are from May 2000 to September 2001, April 2002 to March 2003, July 2007 to February 2009, and May 2011 to September 2011, totalling 54 months. The rest of the months (126) are classified as bullish). Panel C presents the average bear-month returns for the same 37 portfolios, whereas Panel D shows the corresponding bull-month returns (The classification between the bear- and bull-months is based on the sign of the market return in each month. This criterion produces 75 bear months and 105 bull months). The right-most column shows the corresponding statistics for the market portfolio. The portfolio-formation criteria are introduced on the column header line in each panel. In each panel, the first row presents the results for decile portfolios, whereas the second shows the results for the quantile portfolios that are of the same size as the *F*-score-boosted portfolios introduced on the bottom row. The number of constituent stocks in each quantile portfolio equals the number of stocks with the two highest *F*-scores (i.e., either 8 or 9) in the corresponding top-tertile (bottom-tertile) portfolio formed on each value or composite measure (or *ACCR*) and varies over 1-year holding periods. The plain *F*-score portfolio is exceptional in that its constituent stocks (i.e., those with the two highest *F*-scores) are picked from the whole universe of investable stocks without any tertile restrictions. *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage comparable quantile portfolios (i.e., from 1.89 to 3.26% p.m.) and decile portfolios (from 2.15 to 2.99% p.m.).

In the full-sample-period comparison with the corresponding decile portfolios, the *F*-score-boosted quantile portfolios generate higher returns as well as higher SKASRs in all 12 comparable cases. The 4-factor alphas are also higher for *F*-score portfolios in 11 out of 12 cases (The only exception is the combination of S/P and E/P, for which the decile portfolio approach generates a 4-alpha of 8.52 versus 7.66% p.a. obtained using the corresponding *F*-score-boosted approach. However, the *t*-statistic of the latter alpha is higher than that of the former). In most cases, the superiority of *F*-score-boosted portfolios over the comparable quantile portfolios also holds, with the exceptions of the combination of S/P and E/P, and that of S/P and CFO/P, for which the 4-factor alphas of the quantile portfolios formed without *F*-score are slightly higher, but nevertheless less significant than those reported for their *F*-score-boosted counterpart portfolios. For the latter combination, the SKASR is also higher without *F*-score than with it.

One advantage of using *F*-score as a supplementary criterion in value portfolio selection is that it seems to work well coupled with all the 12 portfolio-formation criteria examined. The cross-sectional distribution of average annual returns of *F*-score-boosted portfolios is surprisingly discrete, ranging from 19.00 to 23.98% and averaging 21.92%. By contrast, the corresponding range for the comparable quantile portfolios is from 1.48 to 22.04% p.a. (12.70% p.a., on average), whereas for the decile portfolios, it is from 5.59 to 19.67% p.a. (averaging 13.52% p.a.). In light of these results, and at least for this particular sample data, the applicability of *F*-score seems to extend far beyond what was originally suggested by Piotroski (2000).

Following Piotroski (2000), Bird and Whitaker (2003), and Bird and Casavecchia (2007), we also calculate the hit rates that indicate the proportion of stocks whose returns have been higher than that of the sample average for each quantile portfolio. Table 8 shows that these results are in line with those of Piotroski (2000); for all 12 portfolio-formation criteria, the addition of *F*-score enhances the hit rates. The average hit rate for *F*-score-boosted portfolios is well above 50% (i.e., 56.51%), ranging from 53.77% for E/P to 59.97% for S/P, whereas for the corresponding quantile portfolios formed without *F*-score, the average hit rate is 47.08%, ranging from 36.30% for accruals-to-assets to 53.58% for the combination of S/P and E/P. The hit rate comparisons reinforce the relative added-value differences (stemming from the inclusion of *F*-score) observed on the basis of three performance metrics. The corresponding proportion of stocks with above-average returns for the decile portfolios is 47.45%, on average, ranging from 38.97% for accruals-to-assets to 53.08% for the combination of S/P and E/P. Compared to the hit rates reported in previous studies, our rates are generally somewhat higher than those documented by Piotroski (2000) for U.S. markets and those of Bird and Casavecchia (2007) for pooled European sample data, but they are relatively close to those documented for investment strategies combining value and momentum indicators (see Bird and Casavecchia 2007 for the pan-European evidence and Leivo and Pätäri 2011 for the Finnish evidence).

The other potential pitfall with the use of *F*-score as a supplementary criterion is that it requires a large universe of stocks, within which there must be enough stocks that simultaneously have both a low relative value and high *F*-score. Although the German stock market is among the ten largest national stock market in the world in terms of total capitalization (at the time of writing this, i.e., on the first of July, 2017), the 15-year sample period includes years during which the number of such firms has been questionably low, particularly if an investor would have had restrictions to invest in small-cap firms. In such a case, an investor might have faced difficulties ensuring a sufficient degree of portfolio

Table 8 The average proportions of outperforming stocks in fraction portfolios

	B/P	CFO/P	E/P	S/P	*ACCR*	*MLEV*	B/P and E/P	B/P and S/P	B/P and ROE	B/P and *ACCR*	S/P and CFO/P	S/P and E/P	Plain *F*-score
Decile portfolio	44.43	48.76	49.47	42.98	38.97	44.94	52.13	46.92	51.06	45.84	50.77	53.08	
Quantile portfolio	42.75	49.89	52.12	44.26	36.30	46.22	48.45	44.08	49.88	44.04	53.35	53.58	
F-score-boosted portfolio	55.80	56.33	53.77	59.97	56.49	55.79	55.32	58.12	56.71	54.56	58.13	57.07	53.92

For each of the 37 portfolios examined, the table shows the average proportions (in percentage points and calculated over the 15-year sample period) of stocks whose returns have been higher than that of the sample average return. The portfolio-formation criteria are introduced on the column header line. The first row presents the results for decile portfolios, whereas the second shows the results for the quantile portfolios that are of the same size as the *F*-score-boosted portfolios introduced on the bottom row. The number of constituent stocks in each quantile portfolio equals the number of stocks with the two highest *F*-scores (i.e., either 8 or 9) in the corresponding top-tertile (bottom-tertile) portfolio formed on each value or composite measure (or *ACCR*) and varies over 1-year holding periods. The plain *F*-score portfolio is exceptional in that its constituent stocks (i.e., those with the two highest *F*-scores) are picked from the whole universe of investable stocks without any tertile restrictions. *ACCR* refers to accruals-to-assets, whereas *MLEV* refers to market leverage diversification throughout the sample period if Germany had been the only equity market in which he or she was operating. Although this problem is partially alleviated by lowering the threshold used in determining the high F-score stocks, the possibility of an occasional shortage of investable F-score stocks cannot be ruled out completely, even in a stock market as large as Germany's.[38]

Like most academic peer-group studies, we have not included transaction costs in our analysis because their level is both investor- and trade-specific (see, e.g., Keim and Madhavan 1997; Lewellen 2010). Although their omission causes a small upward bias in the performance metrics of quantile portfolios, recent evidence shows that such a bias is smaller for low-turnover strategies such as those examined in this study (see, e.g., Frazzini et al. 2015; Novy-Marx and Velikov 2016). In addition, many of the stocks in a certain quantile remain in the same quantile after the reformation of portfolios, and in such cases only rebalancing trades rather than the sale or purchase of total stockholdings are needed. On the other hand, the turnover rates would likely be somewhat higher for F-score-boosted portfolios than for plain value portfolios. However, this increasing impact on transaction costs might be at least partially compensated by the documented tendency of the F-score inclusion to shift the market-cap exposure of the portfolios towards larger-cap stocks, as the total relative transaction costs are higher for smaller-cap stocks due to a stronger price impact from implementing trades (e.g., see Chiyachantana et al. 2004). Therefore, the inclusion of transaction costs would only have a marginal impact on comparisons of the relative performance of quantile portfolios, although it would slightly reduce the net profits gained by investors following the examined value strategies.

4 Conclusions

The results show that value anomalies exist in the German stock market during the 2000–2015 period. However, the individual valuation ratios are not the best way to profit from these anomalies, as investors would have benefitted remarkably from combining valuation ratios or using financial statement variables as a basis for a supplementary criterion. According to the results, Piotroski's F-score is particularly useful for the latter purpose. Comparing equal-sized quantile portfolios formed with and without F-score, the performance of the former is better in every pairwise comparison of the 12 comparable cases in terms of both raw returns and the standard Sharpe ratios. The same also holds in terms of the SKASRs, except for the combination portfolio formed on S/P and CFO/P, for which the SKASR is outstandingly higher than the Sharpe ratio, due to its positively skewed and exceptionally leptokurtic return distribution. In terms of the 4-factor alphas, there are two similar exceptions: the combinations of S/P and CFO/P, and S/P and E/P, which generate both the highest SKASRs and the highest 4-factor alphas among the quantile portfolios formed without F-scores.

Compared to the performance of the value portfolios formed without F-scores, the performance of the F-score-boosted quantile portfolios is more even in terms of all three employed performance metrics, implying that adding F-score enhances the efficacy of such portfolio-formation criteria, for which the relative performance is poor without F-score. In this respect, our results suggest that the applicability of F-score is much wider than

[38] E.g., for the sample of Australian stocks, Hyde (2016) lowers the threshold from 8 to 7 to maintain a sufficient number of investable stocks.

originally suggested by Piotroski (2000), who used it for selecting the financially strong firms among the high B/P ones.

The results further show that coupling measures of relative value with *F*-score provides a cushion against declining stock prices, as the average losses during bearish periods are remarkably smaller for *F*-score-boosted portfolios than for the comparable quantile portfolios or the corresponding decile portfolios formed without *F*-scores. Hence, the added-value of *F*-score inclusion is higher during bearish periods.

The use of *F*-score as a supplementary criterion also increases the proportion of stocks that earn higher returns than the stock market average during the subsequent holding period. For this particular sample data, the inclusion of *F*-score besides a relative value measure also tends to increase the average market cap of portfolio firms. A closer examination of the time-series returns of the portfolio formed solely on the basis of high *F*-scores without any relative value criterion reveals that the return correlation between high *F*-score stocks and momentum stocks is very high. To the best of our knowledge, this is the first time when such a strong relationship between high *F*-scores and price momentum indicators has been documented in the financial literature. This finding also provides an interesting topic for further research, as both financial strength and momentum indicators have been documented to work well as another supplementary combination criterion beside a relative value criterion. However, as high *F*-score stocks clearly behave differently to momentum stocks, for example, in bearish stock market conditions, it would be interesting to test whether the combination criteria taking account of all these three style dimensions would work even better for purposes of equity portfolio selection. In addition, the methods employed in this paper could be applied to other stock markets to examine to what extent our results based on the German sample data are generalizable.

Acknowledgements We thank the editor (Cheng-few Lee) and the anonymous referee for their valuable comments and suggestions. We are also grateful to Roman Brückner, Patrick Lehmann, Martin H. Schmidt, and Richard Stehle for making their comprehensive dataset on the German stock markets freely available. In addition, we are indebted to Ville Karell for excellent support and research assistance.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1: The calculation principles for the portfolio-formation criteria

The principles followed in the calculation of the components for the portfolio-formation criteria are as follows:

- Market Value of Equity (P): The stock price(s) multiplied by shares outstanding as end of April of year t.
- Book Value of Equity (B) is *Stockholders' Equity* plus *Deferred Taxes* minus *German-specific tax-deductible reserves* as end of fiscal year $t - 1$.
- Earnings (E) is defined as *Net Income before Extraordinary Items* as end of fiscal year $t - 1$.

Operating Cash Flow (CFO): Following Dissanaike and Lim (2010) and Robin and Wu (2015), CFO is *Earnings* plus *Depreciation and Amortization* minus *Working Capital Accruals* as end of fiscal year $t-1$.[39]

- Sales (S) are from the income statement of year as end of fiscal year $t-1$.
- Accruals-to-assets (*ACCR*): Following Piotroski (2000) and Chen et al. (2017), *ACCR* is [*Net Income before Extraordinary Items* (as end of fiscal year $t-1$) minus *Operating Cash Flow* (as end of fiscal year $t-1$)]/*Total Assets* as end of fiscal year $t-2$.
- Market leverage (*MLEV*) is *Total Assets* as end of fiscal year $t-1$/*Market Value of Equity* as end of April year *t*.
- Return on equity (*ROE*) is *Earnings* as end of fiscal year $t-1$/*Book Value of Equity* as end of fiscal year $t-2$.

Appendix 2

See Table 9.

Table 9 *F*-score variable definitions

Name	Definition	Formula
F1 (*Return on Assets = ROA*)	*Earnings* as end of fiscal year $t-1$/*Total Assets* as end of fiscal year $t-2$	F1 = 1 if *ROA* > 0, otherwise F1 = 0
F2 (Δ*ROA*)	*ROA* as end of fiscal year $t-1$ − *ROA* 1 year prior	F2 = 1 if Δ*ROA* > 0, otherwise F2 = 0
F3 (*Operating Cash Flow = CFO*)	*CFO* as end of fiscal year $t-1$	F3 = 1 if *CFO* > 0, otherwise F3 = 0
F4 (*Accruals*)	*Earnings* as end of fiscal year $t-1$ − *Operating Cash Flow* as end of fiscal year $t-1$	F4 = 1 if *Accruals* < 0, otherwise F4 = 0
F5 (Δ*Gross margin ratio*)	(*Gross Margin/Sales*) as end of fiscal year $t-1$ − (*Gross Margin/Sales*) 1 year prior	F5 = 1 if Δ*Gross margin ratio* > 0, otherwise F5 = 0
F6 (Δ*Asset turnover ratio*)	*Sales* as end of fiscal year $t-1$/*Total Assets* as end of fiscal year $t-2$ − *Sales* as end of fiscal year $t-2$/*Total Assets* as end of fiscal year $t-3$	F6 = 1 if Δ*Asset turnover ratio* > 0, otherwise F6 = 0
F7 (Δ*Leverage*)	(*Total Long-term Debt/Total Assets*) as end of fiscal year $t-1$ − (*Total Long-term Debt/Total Assets*) 1 year prior	F7 = 1 if Δ*Leverage* < 0, otherwise F7 = 0
F8 (Δ*Liquidity*)	(*Current Assets/Current Liabilities*) as end of fiscal year $t-1$ − (*Current Assets/Current Liabilities*) 1 year prior	F8 = 1 if Δ*Liquidity* > 0, otherwise F8 = 0
F9 (Equity offerings = *EQ_OFFER*)	Net stock issues during the fiscal year $t-1$	F9 = 1 if *EQ_OFFER* ≤ 0, otherwise F9 = 0

[39] Working capital accruals (WCA) are calculated following the definition of Dasgupta et al. (2011): WCA = (ΔCurrent Assets (item ACT) − ΔCash (item CH)) − (ΔCurrent Liabilities (item LCT) − ΔShort-term Debt (item DLC) − ΔTax Payable (item TXP)), where Δ represents the annual change.

References

Alles L, Murray L (2010) Non-normality and risk in developing Asian markets. Rev Pac Basin Financ Mark Polic 13:583–605. https://doi.org/10.1142/S0219091510002086

Ang TC (2015) Are firms with negative book equity in financial distress? Rev Pac Basin Financ Mark Polic 18:1–41. https://doi.org/10.1142/S0219091515500162

Artmann S, Finter P, Kempf A (2012a) Determinants of expected stock returns: large sample evidence from the German market. J Bus Finance Account 39:758–784. https://doi.org/10.1111/j.1468-5957.2012.02286.x

Artmann S, Finter P, Kempf A, Koch S, Theissen E (2012b) The cross-section of German stock returns: new data and evidence. Schmalenbach Bus Rev 64:20–43

Asness CS, Moskowitz TJ, Pedersen LH (2013) Value and momentum everywhere. J Finance 68:929–985. https://doi.org/10.1111/jofi.12021

Ball R, Gerakos J, Linnainmaa JT, Nikolaev V (2016) Accruals, cash flows, and operating profitability in the cross section of stock returns. J Financ Econ 121:28–45. https://doi.org/10.1016/j.jfineco.2016.03.002

Barbee WC, Jeong J-G, Mukherji S (2008) Relations between portfolio returns and market multiples. Glob Finance J 19:1–10. https://doi.org/10.1016/j.gfj.2008.02.001

Barroso P, Santa-Clara P (2015) Momentum has its moments. J Financ Econ 116:111–120. https://doi.org/10.1016/j.jfineco.2014.11.010

Barth ME, Hutton AP (2004) Analyst earnings forecast revisions and the pricing of accruals. Rev Account Stud 9:59–96. https://doi.org/10.1023/B:RAST.0000013629.59222.df

Bartov E, Kim M (2004) Risk mispricing and value investing. Rev Quant Finance Account 23:353–376. https://doi.org/10.1023/B:REQU.0000049321.34133.95

Beaver WH (2002) Perspectives on recent capital market research. Account Rev 77:453–474. https://doi.org/10.2308/accr.2002.77.2.453

Bird R, Casavecchia L (2007) Value enhancement using momentum indicators: the European experience. Int J Manag Finance 3:229–262. https://doi.org/10.1108/17439130710756907

Bird R, Whitaker J (2003) The performance of value and momentum investment portfolios: recent experience in the major European markets. J Asset Manag 4:221–246. https://doi.org/10.1057/palgrave.jam.2240105

Brown SJ, Du DY, Rhee SG, Zhang L (2008a) The returns to value and momentum in Asian markets. Emerg Mark Rev 9:79–88. https://doi.org/10.1016/j.ememar.2008.02.001

Brown SJ, Lajbcygier P, Li B (2008b) Going negative: what to do with negative book equity stocks. J Portf Manag 35:95–102. https://doi.org/10.3905/JPM.2008.35.1.95

Brückner R (2013) Important characteristics, weaknesses and errors in German equity data from Thomson Reuters Datastream and their implications for the size effect. Working paper (Humboldt University of Berlin). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2243816. Accessed 18 May 2016. https://doi.org/10.2139/ssrn.2243816

Brückner R, Lehmann P, Stehle R (2012) In Germany the CAPM is alive and well. Working paper, (Humboldt University of Berlin). http://ssrn.com/abstract=2161847. Accessed 17 May 2016

Brückner R, Lehmann P, Schmidt MH, Stehle R (2015) Non-U.S. multi-factor data sets should be used with caution. Working paper (Humboldt University of Berlin). http://ssrn.com/abstract=2390063. Accessed 17 May 2016

Cakici N, Fabozzi FJ, Tan S (2013) Size, value, and momentum in emerging market stock returns. Emerg Mark Rev 16:46–65. https://doi.org/10.1016/j.ememar.2013.03.001

Cakici N, Topyan K, Wang CJ (2014) Cross-sectional return predictability in Taiwan stock exchange: an empirical investigation. Rev Pac Basin Financ Mark Polic 17:1–44. https://doi.org/10.1142/S0219091514500106

Capaul C, Rowley I, Sharpe WF (1993) International value and growth stock returns. Financ Anal J 49:27–36. https://doi.org/10.2469/faj.v49.n1.27

Carhart MM (1997) On persistence in mutual fund performance. J Finance 52:57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chan L, Lakonishok J (2004) Value and growth investing: review and update. Financ Anal J 60:71–86. https://doi.org/10.2469/faj.v60.n1.2593

Chen H, Dong W, Han H, Zhou N (2017) A comprehensive and quantitative internal control index: construction, validation, and impact. Rev Quant Finance Account 49:337–377. https://doi.org/10.1007/s11156-016-0593-x

Chiyachantana CN, Jain PK, Jiang C, Wood RA (2004) International evidence on institutional trading behavior and price impact. J Finance 59:869–898. https://doi.org/10.1111/j.1540-6261.2004.00651.x

Clinch G, Fuller D, Govendir B, Wells P (2012) The accrual anomaly: Australian evidence. Account Finance 52:377–394. https://doi.org/10.1111/j.1467-629X.2010.00380.x
Collins D, Hribar P (2000) Earnings-based and accrual-based market anomalies: one effect or two? J Account Econ 29:101–123. https://doi.org/10.1016/S0165-4101(00)00015-X
Cornish EA, Fisher RA (1937) Moments and cumulants in the specification of distributions. Rev Inst Int Stat 5:307320. https://doi.org/10.2307/1400905
Damodaran A (2012) Investment valuation. Wiley, Hoboken
Dasgupta S, Noe TH, Wang Z (2011) Where did all the dollars go? The effect of cash flows on capital and asset structure. J Financ Quant Anal 46:1259–1294. https://doi.org/10.1017/S0022109011000512
Dechow P, Richardson S, Sloan R (2008) The persistence and pricing of the cash component of earnings. J Account Res 46:537–566. https://doi.org/10.1111/j.1475-679X.2008.00283.x
Dechow P, Khimich NV, Sloan R (2011) The accrual anomaly. In: Zacks L (ed) The handbook of equity market anomalies: translating market inefficiencies into effective investment strategies. Wiley, Hoboken, pp 23–62
Desai H, Rajgopal S, Venkatachalam M (2004) Value-glamour and accruals mispricing: one anomaly or two? Account Rev 79:355–385. https://doi.org/10.2308/accr.2004.79.2.355
Dhatt MS, Kim YJ, Mukherji S (2004) Can composite value measures enhance portfolio performance? J Invest 13:42–48. https://doi.org/10.3905/joi.2004.450755
Dissanaike G, Lim K-H (2010) The sophisticated and the simple: the profitability of contrarian strategies. Eur Financ Manag 16:229–255. https://doi.org/10.1111/j.1468-036X.2008.00466.x
Dopuch N, Seethamraju C, Xu W (2010) The pricing of accruals for profit and loss firms. Rev Quant Finance Account 34:505–516. https://doi.org/10.1007/s11156-009-0144-9
Edwards S, Biscarri JG, Pérez de Gracia F (2003) Stock market cycles, financial liberalization and volatility. J Int Money Finance 22:925–955. https://doi.org/10.1016/j.jimonfin.2003.09.011
Fama EF, French KR (1992) The cross-section of expected stock returns. J Finance 47:427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x
Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33:3–56. https://doi.org/10.1016/0304-405X(93)90023-5
Fama EF, French KR (1998) Value versus growth: the international evidence. J Finance 53:1975–1999. https://doi.org/10.1111/0022-1082.00080
Fama EF, French KR (2006) The value premium and the CAPM. J Finance 61:2163–2185. https://doi.org/10.1111/j.1540-6261.2006.01054.x
Fama EF, French KR (2008) Dissecting anomalies. J Finance 63:1653–1678. https://doi.org/10.1111/j.1540-6261.2008.01371.x
Fama EF, French KR (2011) Why use book value to sort stocks? Fama/French Forum. https://famafrench.dimensional.com/questions-answers/qa-why-use-book-value-to-sort-stocks.aspx. Accessed 10 Aug 2017
Fama EF, French KR (2012) Size, value, and momentum in international stock returns. J Financ Econ 105:457–472. https://doi.org/10.1016/j.jfineco.2012.05.011
Favre L, Galéano J-A (2002) Mean-modified value-at-risk optimization with hedge funds. J Altern Invest 5:21–25. https://doi.org/10.3905/jai.2002.319052
Feunou B, Jahan-Parvar MR, Tédongap R (2013) Modeling market downside volatility. Rev Finance 17:443–481. https://doi.org/10.1093/rof/rfr024
Fisher G, Shah R, Titman S (2016) Combining value and momentum. J Invest Manag 14:33–48
Frazzini A, Israel R, Moskowitz T (2015) Trading costs of asset pricing anomalies. Working paper (University of Chicago, Illinois). http://faculty.chicagobooth.edu/tobias.moskowitz/research/Trading%20Cost%20of%20Asset%20Pricing%20Anomalies_full.pdf. Accessed 1 June 2016
Fuller KP, Goldstein MA (2011) Do dividends matter more in declining markets? J Corp Finance 17:457–473. https://doi.org/10.1016/j.jcorpfin.2011.01.001
Gharghori P, Stryjkowski S, Veeraraghavan M (2013) Value versus growth: Australian evidence. Account Finance 53:393–417. https://doi.org/10.1111/j.1467-629X.2012.00474.x
Gordon MJ, Shapiro E (1956) Capital equipment analysis: the required rate of profit. Manag Sci 3:102–110. https://doi.org/10.1287/mnsc.3.1.102
Gray WR, Vogel J (2012) Analyzing valuation measures: a performance horse-race over the past 40 years. J Portf Manag 39:112–121. https://doi.org/10.3905/jpm.2012.39.1.112
Haugen R, Baker N (1996) Commonality in the determinants of expected stock returns. J Financ Econ 41:401–439. https://doi.org/10.1016/0304-405X(95)00868-F
Homm U, Pigorsch C (2012) Beyond the Sharpe ratio: an application of the Aumann–Serrano index to performance measurement. J Bank Finance 36:2274–2284. https://doi.org/10.1016/j.jbankfin.2012.04.005

Hou K, Xue C, Zhang L (2017) A comparison of new factor models. Working paper (Fisher College of Business, Ohio). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2520929. Accessed 10 Aug 2017
Hyde CE (2016) The Piotroski F-score: evidence from Australia. Account Finance. https://doi.org/10.1111/acfi.12216
Ince OS, Porter RB (2006) Individual equity return data from Thomson Datastream: handle with care! J Financ Res 29:463–479. https://doi.org/10.1111/j.1475-6803.2006.00189.x
Israel R, Moskowitz TJ (2013) The role of shorting, firm size, and time on market anomalies. J Financ Econ 108:275–301. https://doi.org/10.1016/j.jfineco.2012.11.005
Israelsen CL (2005) A refinement to the Sharpe ratio and information ratio. J Asset Manag 5:423–427. https://doi.org/10.1057/palgrave.jam.2240158
Keim DB, Madhavan A (1997) Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. J Financ Econ 46:265–292. https://doi.org/10.1016/S0304-405X(97)00031-7
Kim JB, Lipka R, Sami H (2012) Portfolio performance and accounting measures of earnings: an alternative look at usefulness. Rev Quant Financ Account 38:87–107. https://doi.org/10.1007/s11156-010-0220-1
Lakonishok J, Shleifer A, Vishny R (1994) Contrarian investment, extrapolation, and risk. J Finance 49:1541–1578. https://doi.org/10.1111/j.1540-6261.1994.tb04772.x
Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the Sharpe ratio. J Empir Finance 15:850–859. https://doi.org/10.1016/j.jempfin.2008.03.002
Lee HL, Lee H (2015) Effect of information disclosure and transparency ranking system on mispricing of accruals of Taiwanese firms. Rev Quant Financ Acc 44:445–471. https://doi.org/10.1007/s11156-013-0413-5
Lee CF, Su JB (2012) Alternative statistical distributions for estimating value-at-risk: theory and evidence. Rev Quant Financ Acc 39:309–331. https://doi.org/10.1007/s11156-011-0256-x
Leibowitz MA (1999) Market-to-book ratios and positive and negative returns on equity. J Financ Statement Anal 4:21–30. https://doi.org/10.2469/dig.v29.n3.512
Leivo TH (2012) Combining value and momentum indicators in varying stock market conditions: the Finnish evidence. Rev Account Finance 11:400–447. https://doi.org/10.1108/14757701211279187
Leivo TH, Pätäri EJ (2011) Enhancement of value portfolio performance using momentum and the long-short strategy: the Finnish evidence. J Asset Manag 11:401–416. https://doi.org/10.1057/jam.2009.38
Leshem R, Goldberg LR, Cummings A (2016) Optimizing value. J Portf Manag 42:77–89. https://doi.org/10.3905/jpm.2016.42.2.077
Lev B, Nissim D (2006) The persistence of the accruals anomaly. Contemp Account Res 23:193–226. https://doi.org/10.1506/C6WA-Y05N-0038-CXTB
Lewellen J (2010) Accounting anomalies and fundamental analysis: an alternative view. J Account Econ 50:455–466. https://doi.org/10.1016/j.jacceco.2010.09.007
Loughran T, Wellman J (2011) New evidence on the relation between the enterprise multiple and average stock returns. J Financ Quant Anal 46:1629–1650. https://doi.org/10.1017/S0022109011000445
Newey WK, West KD (1987) A simple semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55:703–708. https://doi.org/10.2307/1913610
Novy-Marx R (2012) Is momentum really momentum? J Financ Econ 103:429–453. https://doi.org/10.1016/j.jfineco.2011.05.003
Novy-Marx R (2013) The other side of value: the gross profitability premium. J Financ Econ 108:1–28. https://doi.org/10.1016/j.jfineco.2013.01.003
Novy-Marx R (2014) Quality investing. Working paper (University of Rochester, New York). http://rnm.simon.rochester.edu/research/QDoVI.pdf. Accessed 10 Aug 2017
Novy-Marx R, Velikov M (2016) A taxonomy of anomalies and their trading costs. Rev Financ Stud 29:104–147. https://doi.org/10.1093/rfs/hhv063
Ou J, Penman S (1989) Financial statement analysis and the prediction of stock returns. J Account Econ 11:295–329. https://doi.org/10.1016/0165-4101(89)90017-7
Pätäri EJ (2011) Does the risk-adjustment method matter at all in hedge fund rankings? Int Res J Finance Econ 6:69–99
Pätäri EJ, Leivo TH (2017) A closer look at value premium: literature review and synthesis. J Econ Surv 31:79–168. https://doi.org/10.1111/joes.12133
Pätäri EJ, Tolvanen J (2009) Chasing performance persistence of hedge funds—comparative analysis of evaluation techniques. J Deriv Hedge Funds 15:223–240. https://doi.org/10.1057/jdhf.2009.11
Pätäri EJ, Leivo TH, Honkapuro JVS (2010) Enhancement of value portfolio performance using data envelopment analysis. Stud Econ Finance 27:223246. https://doi.org/10.1108/10867371011060036
Pätäri EJ, Leivo TH, Honkapuro JVS (2012) Enhancement of equity portfolio performance using data envelopment analysis. Eur J Oper Res 220:786–797. https://doi.org/10.1016/j.ejor.2012.02.006

Pätäri EJ, Karell V, Luukka P (2016) Can size-, industry-, and leverage-adjustment of valuation ratios benefit the value investor? Int J Bus Innov Res 11:76–109. https://doi.org/10.1504/IJBIR.2016.077611
Pätäri EJ, Karell V, Luukka P, Yeomans JS (2017a) Comparison of the multicriteria decision-making methods for equity portfolio selection: the US evidence. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2017.08.001
Pätäri EJ, Karell V, Luukka P, Yeomans JS (2017b) The dirty dozen of valuation ratios: Is one better than the other? J Invest Manag (in press)
Penman S (1991) An evaluation of accounting rate-of-return. J Account Audit Finance 6:233–255. https://doi.org/10.1177/0148558X9100600204
Penman S, Reggiani F (2013) Returns to buying earnings and book value: accounting for growth and risk. Rev Account Stud 18:1021–1049. https://doi.org/10.1007/s11142-013-9226-y
Piotroski J (2000) Value investing: the use of historical financial statement information to separate winners from losers. J Account Res 38:1–41. https://doi.org/10.2307/2672906
Piotroski J, So EC (2012) Identifying expectation errors in value/glamour strategies: a fundamental analysis approach. Rev Financ Stud 25:2841–2875. https://doi.org/10.1093/rfs/hhs061
Richardson S, Tuna I, Wysocki P (2010) Accounting anomalies and fundamental analysis: a review of recent research advances. J Account Econ 50:410–454. https://doi.org/10.1016/j.jacceco.2010.09.008
Robin A, Wu Q (2015) Firm growth and the pricing of discretionary accruals. Rev Quant Financ Account 45:561–590. https://doi.org/10.1007/s11156-014-0447-3
Sharpe WF (1966) Mutual fund performance. J Bus 39:119–138. https://doi.org/10.1086/294846
Simlai PE (2016) Time-varying risk, mispricing attributes, and the accrual premium. Int Rev Financ Anal 48:150–161. https://doi.org/10.1016/j.irfa.2016.09.014
Sloan RG (1996) Do stock prices fully reflect information in accruals and cash flows about future earnings? Account Rev 71:289–315. https://doi.org/10.2469/dig.v27.n1.5
Stehle R (1997) Der Size-Effekt am deutschen Aktienmarkt. Zeitschrift für Bankrecht und Bankwirtschaft 9:237–260
Stehle R, Schmidt MH (2015) Returns on German stocks 1954 to 2013. Credit Cap Mark 48:427–476. https://doi.org/10.3790/ccm.48.3.427
Stehle R, Brückner R, Lehmann P, Schmidt MH (2015) Fama–French factors for Germany (Dataset). Humboldt University of Berlin. https://www.wiwi.hu-berlin.de/de/professuren/bwl/bb/data/fama-french-factors-germany/fama-french-factors-for-germany. Accessed 2 May 2016
Taylor S, Wong L (2012) Robust anomalies? A close look at accrual-based trading strategy returns. Account Finance 52:573–603. https://doi.org/10.1111/j.1467-629X.2011.00408.x
Theodossiou P, Savva C (2016) Skewness and the relation between risk and return. Manag Sci 62:1598–1609. https://doi.org/10.1287/mnsc.2015.2201
Vitols S (2001) Frankfurt's Neuer Market and the IPO explosion: is Germany on the road to Silicon Valley? Econ Soc 30:553–564. https://doi.org/10.1080/03085140120089090
Wilcox JW (1984) The P/B-ROE valuation model. Financ Anal J 40:58–66. https://doi.org/10.2469/faj.v40.n1.58
Wilcox JW, Philips TK (2005) The P/B-ROE valuation model revisited. J Portf Manag 31:56–66. https://doi.org/10.3905/jpm.2005.570151
Zakamouline V, Koekebakker S (2009) Portfolio performance evaluation with generalized Sharpe ratios: beyond the mean and variance. J Bank Finance 33:1242–1254. https://doi.org/10.1016/j.jbankfin.2009.01.005