Journal Of Investment Management, Vol. 16, No. 2, (2018), pp. 65–98

JOIM
www.joim.com

THE DIRTY DOZEN OF VALUATION RATIOS: IS ONE BETTER THAN ANOTHER?

Eero Pätäri[a,*], *Ville Karell*[a], *Pasi Luukka*[a] *and Julian Scott Yeomans*[b]

This paper compares the efficacy of both traditional valuation ratios and an extensive set of related combination criteria in identifying the future best-performing stocks for a comprehensive U.S. sample over the period 1971–2013. Value portfolios formed on different criteria have remarkably different exposures to style factors. We find evidence of strong relative efficacy of three enterprise value multiples (EBIT/EV, EBITDA/EV, and S/EV). Particularly, the evidence for the unique characteristics of S/EV contributes to the existing literature. The defensive characteristic of high dividend yield is pervasive both as a stand-alone criterion and as a combination sub-criterion.

1 Introduction

The extensive literature on value anomalies has shown that value stocks tend to outperform the market portfolio most of the time (see, e.g., Chan and Lakonishok, 2004; Fama and French, 2006b, 2012; Cakici *et al.*, 2013). Evidence has not only shown that the value anomalies in stock markets are a global phenomenon, but also that the relative efficacy of different valuation criteria varies across both the stock markets and the sample periods examined (for a comprehensive literature review on value anomalies, see Pätäri and Leivo, 2017). For example, Fama and French (1998) compared the value portfolio returns and value premiums obtained by using four different portfolio-formation criteria (i.e., *book*-to-*price* (B/P), *cash flow*-to-*price* (CF/P), *earnings yield* (E/P), and *dividend yield* (D/P)) in 13 major stock markets. According to their results, the classification criteria resulting in the highest value premium and/or the highest value portfolio return for the period 1975–1995 varied across countries.

Within the U.S. markets, Davis (1994) documented the highest value quintile return and the highest value premium for the E/P criterion for the period 1940–1963. However, Lakonishok *et al.*

[a]School of Business and Management, Lappeenranta University of Technology, P. O. Box 20, FIN-53851 Lappeenranta, Finland.
[b]Schulich School of Business, York University, Toronto, ON, M3J 1P3, Canada.
*Corresponding author. Tel: +358 (0)40 833 6907. E-mail: eero.patari@lut.fi

(1994) found that the CF/P criterion resulted in both the highest value decile portfolio return and the highest value premium for the period 1968–1990. Desai *et al.* (2004) also reported the highest value decile return for the same criterion, but found that the highest value premium was generated by the B/P criterion in their comparison of the efficacy of four individual valuation ratios (i.e., B/P, CF/P, E/P, and *operating cash flow*-to-*price* (CFO/P)) for the period 1973–1997. By contrast, Israel and Moskowitz (2013) reported the highest value decile return for the B/P criterion, whereas the highest value premium for the period 1951–2011 was documented for the CF/P criterion, consistent with the results of Lakonishok *et al.* (1994) for the earlier period. In addition, Li *et al.* (2009) documented the highest decile return and the highest value premium for the E/P criterion over 1963–2006 in an efficacy comparison of B/P, CF/P, and E/P whereas, according to Loughran and Wellman (2011), the E/P criterion was the second worst in the corresponding comparison of B/P, D/P, E/P, and *earnings before interest, taxes, depreciation and amortization*-to-*enterprise value* (EBITDA/EV) in the 1963–2009 sample period. Recent literature has also provided evidence that enterprise value (EV) multiples may add value to the portfolio selection. In an efficacy comparison of five individual valuation ratios (B/P, E/P, EBITDA/EV, *free cash flow*-to-*enterprise value* (FCF/EV), and *gross profit*-to-*enterprise value* (GP/EV)), Gray and Vogel (2012) documented the highest value premium, as well as the highest value quintile return, for the EBITDA/EV criterion based on equal-weighted returns. However, in terms of value-weighted quintile returns, the GP/EV criterion was the best. In terms of value-weighted decile returns, Loughran and Wellman (2011) found the highest value premium for the EBITDA/EV criterion, whereas the B/P criterion was the best based on equal-weighted returns.

The above-described diversity of the results on the relative performance of different valuation ratios in the U.S. markets can be explained, for the most part, by differences in the sample selection principles, sample periods, and methodologies employed. For an example of the first type of explanation, one valuation ratio may perform better among small-cap stocks, whereas the other may have a better discriminatory power among large-cap stocks. As an example of the second type of explanation, the recent results of McLean and Pontiff (2016) showed a remarkable post-publication decline in anomalous returns (see also the theory of gradual decay of anomalies introduced by Jones and Pomorski, 2017). With respect to methodological issues, the list of sources of differences is endless; for example, the different treatment practices of companies with negative earnings present one potential explanation (among many others) for the diversity of the earlier results. Although negative E/P stocks have been excluded in the majority of the related studies, their exclusion can create a sample selection bias in two ways. First, if firms with negative income are excluded from the sample, many potential turnaround cases that could be among the best-performing stocks in the future are also excluded.[1] Second, if firms with negative income are excluded only when using earnings or cash flow multiples as a valuation criterion, but not when using other valuation criteria (as is done in the majority of value anomaly studies), the comparability of the results decreases because the samples are not identical. For example, when following such an exclusion policy, a higher value premium or better value quantile portfolio performance based on B/P ratios than based on E/P ratios could be explained by the fact that high (low) B/P firms with negative earnings have, on average, performed well (badly) enough to add value for the value investor. The same kind of dilemma also holds for other earnings or cash flow multiples, such as all variants of CF/P, and *earnings before interest and taxes*-to-*enterprise value* (EBIT/EV), EBITDA/EV, or FCF/EV. For the above-mentioned reasons, we do not exclude negative earnings or cash flow multiples from our samples, unlike the majority of previous studies.

This paper contributes to the current literature on pricing anomalies in several ways. First, we test and compare the discriminatory power of portfolio-formation criteria formed on a larger number of individual valuation ratios than examined in any of the earlier studies. Our analysis includes 12 different individual valuation ratios, of which four are EV multiples. To the best of our knowledge, one of these four (i.e., *sales*-to-*enterprise value* (S/EV)) has not been examined before in this context,[2] and the other three very seldom, although the results of few such studies have been encouraging (see, e.g., Loughran and Wellman, 2011; Gray and Vogel, 2012; Walkshäusl and Lobe, 2015; Pätäri *et al.*, 2016). Because the evidence for a good discriminatory power of EV multiples in detecting under- and/or overvalued stocks is quite recent, they may offer higher anomalous returns than price-based multiples for which the documented history of anomalous returns is much longer (see McLean and Pontiff, 2016; Pätäri and Leivo, 2017). In addition, leverage is only implicitly included in price-based multiples,[3] whereas it has direct influence on EV multiples.

We also test the efficacy of the *sales*-to-*price* (S/P) criterion that has not been included in the recent comparative U.S. studies, although the results of few such studies have shown its high efficacy (see, e.g., Dhatt *et al.*, 2004; Barbee *et al.*, 2008). Although neither the financial literature nor the investment community is unanimous in the applicability of sales multiples for valuation purposes, their use is often motivated by their stability when compared with other valuation multiples, or by the fact that sales are relatively difficult to manipulate, at least in comparison with earnings and book values (see, e.g., Damodaran, 2012). The biggest disadvantage of using sales multiples is that if a firm generates high sales growth while simultaneously losing significant amounts of money, S/P could erroneously indicate a low relative valuation for such a firm (see, e.g., Penman, 2013). Sales can also be increased by increasing debt, which in most cases increases S/P.[4] However, the S/P ratio does not reveal the degree of leverage with which the sales have been generated, although leverage certainly makes a difference to the risks of the firms being compared. For this reason, the S/EV ratio should be more informative than S/P.

We also combine these 12 valuation criteria to find out whether the performance of value portfolios and/or the value premium stemming from combining value indicators can be enhanced from that which is generated by the best single selection criteria. Although the potential benefits of combining individual valuation ratios have been foreseen in several recent academic papers (Israel and Moskowitz, 2013; Asness *et al.*, 2015), the topic has not been examined extensively prior to our paper.

In addition, we check the robustness of our results to varying sample selection criteria. Moreover, we compare the relative efficacy of four variants of CF/P ratios with one another, which has not been done in previous literature on value anomalies. We also evaluate the relative performance of portfolios formed on each selection criterion from several different viewpoints. We use raw returns, the Sharpe ratios, as well as multi-factor alphas for this purpose. In addition, we 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 to find out whether the differences in the sensitivity to rising or declining stock market returns explain the outperformance (underperformance) of the best (worst) portfolios.

For a comprehensive sample data of U.S. stocks over a 42-year period, our results show that there are remarkable differences in the characteristics between the valuation ratios. We find evidence of strong relative efficacy of three EV multiples (EBIT/EV, EBITDA/EV, and S/EV) in comparison with the most commonly used price-based multiples, such as B/P and E/P. In particular, the evidence for the unique characteristics of S/EV contributes to the existing literature. We also show that the dividend yield criterion has been particularly useful for risk reduction purposes to the extent that, in terms of total risk-adjusted returns, the highest D/P decile portfolio is the best among all the decile portfolios formed on the 12 valuation ratios. The low-risk characteristic of the top-decile D/P portfolio also extends to the combination criteria in which D/P is included as one sub-criterion.

The rest of the paper is organized as follows. The first section describes the data and the variables, while the second section explains the methodology employed. The third section introduces the empirical results, whereas the fourth section discusses their robustness. The fifth section summarizes the main findings and also discusses the limitations of the study. The final section concludes.

2 Data and variables

The sample data consist of non-financial firms on the NYSE, AMEX, and NASDAQ with reliable data from the Center for Research in Security Prices (CRSP) and Compustat. The sample comprises only firms with ordinary common equity on CRSP (share code 10 or 11). Adjustments of returns for dividends, splits and capitalization issues are made appropriately. To avoid survivorship bias, CRSP delisting returns are incorporated according to Beaver *et al.* (2007) with the exception that we recalculated new delisting return estimates for firms with missing delisting returns on CRSP for each stock exchange based on the available and traceable delisting returns over the period 1971–2013.[5]

To rank the stocks into decile portfolios, we use 12 different valuation ratios of which eight are price-based (i.e., B/P, CF1/P, CF2/P, CF3/P, CFO/P, D/P, E/P, and S/P) and four are EV-based (i.e., EBIT/EV, EBITDA/EV, FCF/EV, and S/EV). The components of valuation ratios are calculated according to the following principles:

- Market Value of Equity (ME): ME is the stock price multiplied by shares outstanding from the CRSP monthly file and obtained as end of April of year *t* throughout the paper.
- Enterprise Value (EV): Following Loughran and Wellman (2011), EV is computed as Market Value of Equity (ME) plus Short-term Debt (Compustat data item DLC) plus Long-term Debt (item DLTT) plus Preferred Stock Value (item PSTKRV) minus Cash and Short-term Investments (item CHE).
- Earnings (E): Similar to Fama and French (2001), we calculate Earnings (E) as Income Before Extraordinary Items (item IB) minus Preferred Dividends (item DVP) plus Income Statement Deferred Taxes (item TXDI).
- Book Value of Equity (BE): In line with Novy-Marx (2013), book equity is Stockholders' Equity plus Deferred Taxes minus Preferred Stock (Stockholders' Equity is as given in Compustat (SEQ) if available, or else Common Equity plus the Carrying Value of Preferred Stock (CEQ + PSTX) if available, or else Total Assets minus Total Liabilities (AT – LT), when available. Deferred Taxes is Deferred Taxes and Investment Tax Credits (TXDITC)

if available, or else Deferred Taxes and/or Investment Tax Credit (TXDB and/or ITCB). Preferred Stock is Redemption Value (PSTKR) if available, or else Liquidating Value (PSTKL) if available, or else Par Value (PSTK)).

- Sales (S) (item SALE).
- Common Dividends (D) (item DVC): Common dividends paid from the latest firm-specific fiscal year preceding the first of January of the portfolio-formation year.
- Cash Flow 1 (CF1): Following Israel and Moskowitz (2013) and Hou *et al.* (2015), we compute CF1 as Income Before Extraordinary Items (item IB) plus Equity's Share of Depreciation[6] plus Income Statement Deferred Taxes (item TXDI).
- Cash Flow 2 (CF2): Similar to Dhatt *et al.* (2004), we compute CF2 as Earnings Per Share (Diluted) excluding Extraordinary Items (item EPSFX) multiplied by Common Shares Outstanding (item CSHO) plus Income Statement Deferred Taxes (item TXDI) minus Preferred Dividends (item DVP) plus Depreciation and Amortization (item DP).
- Cash Flow 3 (CF3): In line with Hou *et al.* (2011), we compute CF3 as Net Income (item NI) plus Depreciation and Amortization (item DP) plus Income Statement Deferred Taxes (item TXDI).
- Operating Cash Flow (CFO): Following Strydom *et al.* (2014), we compute CFO as Earnings (E) minus Accruals.[7]
- Free Cash Flow (FCF): Similar to Novy-Marx (2013), we calculate FCF as Net Income (item NI) plus Depreciation and Amortization (item DP) minus Working Capital Change (Working Capital (item WCAP) minus prior year's Working Capital) minus Capital Expenditures (item CAPX).
- Earnings Before Interest, Taxes, Depreciation and Amortization (item EBITDA).
- Earnings Before Interest and Taxes (item EBIT).

The accounting-based variables required for the calculation of valuation ratios are collected from the Compustat database, covering 43 years from 1969 to 2011. To be included in the final sample, the firms must have all the information available for the calculation of all 12 single selection criteria being examined in each portfolio-formation point. Although this prerequisite reduced the number of otherwise usable firm-year observations, we require this in order to maintain the best possible comparability of the results based on different selection criteria.[8] For each valuation ratio, we checked, case-by-case, those values that were outside the range of 2.5 percentiles and 97.5 percentiles, and removed them if they were most probably products of database errors. The time-series returns for each stock were also checked accordingly to be able to correct extremely high (low) returns stemming from database error (the most common reason for such cases was the omission of either the reverse split in case of extremely high returns or split in case of extremely low returns). To alleviate backfill bias, in accordance with Fama and French (1993), firms are required to have 2 years of Compustat data before entering our sample. For example, firms included in the sample at the beginning of the sample period (i.e., at the end of April 1971) had to have financial statement data for the fiscal years ending in 1969 and 1970 available in Compustat.[9] In addition, only firms with fiscal year duration of 12 months are included in the final sample. Consistent with the existing literature (see, e.g., Fama and French, 2008; Loughran and Wellman, 2011), we also exclude such firm-year observations for which the book value of equity is negative. The same practice is also followed in cases of negative enterprise value. As in Jegadeesh and Titman (2001), Chan and Lakonishok (2004), and Avramov *et al.* (2007), we also exclude firms for which market capitalization at the beginning of each 1-year holding period is below the bottom NYSE decile breakpoint.[10] After applying all the exclusion criteria, the final sample contains 48,390 firm-year observations with complete data for each of the 12 valuation ratios over the 1971–2013 sample period.

3 Methodology

3.1 Portfolio-formation principles

We construct equal-weighted decile portfolios at the end of April of year t and hold these for one year from May of year t through April of year $t+1$. Like Fama and French (1992, 1993) and Davis *et al.* (2000), we use year $t-1$ financial statement data as constituents of valuation ratios, as there may be a 4-month lag in the publication of financial statement after the end of the fiscal year (see Fama and French, 1992). Following Lakonishok *et al.* (1994), Desai *et al.* (2004), and Fisher *et al.* (2016), we use the latest available information about the market values of equity and enterprise values in portfolio formation (i.e., information at the end of April of year t). In addition to the comparison of 12 individual valuation ratios, we also form 2-, 3-, and 4-combinations of them by calculating the corresponding median-scaled composite measures for each combination.[11] Analogous to Dhatt *et al.* (2004), we first standardize each individual valuation ratio for a firm in a particular year by its cross-sectional median at the same time point and then compute the average of these median-scaled scores for each combination. For example, if B/P for a firm at some point is 0.9, and the median B/P for the sample is 0.6 at the same time point, then the median-scaled B/P for that firm is 1.5.[12] The stocks with the highest valuation ratios or the highest median-scaled composite measures are then put into the top decile, whereas the bottom-deciles consist of the stocks with the lowest ratios or median-scaled measures. The weight changes of the stocks stemming from their return differences within the holding periods are taken into account in the calculation of monthly decile portfolio returns.[13] As a result, we get monthly returns for each deciles over a 42-year investment period.

3.2 Test procedures for performance comparisons

The performance of decile portfolios is primarily evaluated based on the average return and the Sharpe ratio (Sharpe, 1966).[14] To evaluate whether the potential abnormal returns are explained by common explanatory factors, we also calculate the multi-factor alphas and the corresponding factor slopes for each decile portfolio formed on all individual valuation ratios, as well as for those 2-, 3-, and 4-combination portfolios that perform best in terms of either raw returns or the Sharpe ratio.[15] The regression equation is as follows (all the factor returns or, in case of the volatility factor, their constituents were downloaded from French's 2015 data library, available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html):

$$\begin{aligned} r_{it} - r_{ft} = \alpha_i + b_i(r_{mt} - r_{ft}) + s_i SMB_t \\ + h_i HML_t + m_i WML_t \\ + v_i SMV_t + \varepsilon_{it} \end{aligned} \quad (1)$$

where r_{it} = the return of a portfolio

r_{ft} = the risk-free rate of return

α_i = the 5-factor alpha (the abnormal return over and above what might be expected based on the 5-factor model employed)

r_{mt} = the stock market return[16]

SMB_t = the return of the size factor (i.e., the return difference between small- and large-cap portfolios)

HML_t = the return of the book-to-market (B/P) factor (i.e., the return difference between high- and low-B/P portfolios)

WML_t = the return of the momentum factor (i.e., the return difference between winner and loser stock portfolios)

SMV_t = the return of the volatility factor (i.e., the return difference between low- and high-volatility (stable and volatile) portfolios)

b_i, s_i, h_i, m_i and v_i are factor sensitivities to stock market, SMB, HML, WML, and SMV factors, respectively.

ε_{it} = the residual term.

Before ending up adopting this specific 5-factor model, we tested several other factor models to find the one with the best fit for our purpose (the five alternative pricing models were the Fama–French 3- and 5-factor models (see Fama and French, 1993, 2015, respectively), the 4-factor model that is the reduced version of the Fama–French 5-factor model without the HML factor,[17] the Carhart (1997) 4-factor model, and the 6-factor model (introduced in Fama and French, 2016) in which the momentum factor is added to the previously-mentioned 5-factor model). We first chose the Carhart 4-factor model for our analysis because many of the decile portfolios being evaluated had significant momentum exposures despite the facts that momentum was not directly used as a portfolio formation criteria, and that the inclusion of the momentum factor improved the average adjusted R-squareds of the regression runs. We ended up preferring the Carhart 4-factor model instead of the 6-factor model because, for the sample data employed, the inclusion of the RMW[18] factor in pricing models generated the U-shaped relation between decile alphas and corresponding portfolios.[19] This is sharply inconsistent with the results based on total risk-based performance measures, as well as those based on the Carhart 4-factor alphas, both of which indicated, on average, monotonically decreasing performance when moving from top portfolios to bottom portfolios. Motivated by the findings of Clarke *et al.* (2010, 2014) according to whom a volatility-based factor is among the most important in performance measurement of equity portfolios, we added a volatility factor as the fifth factor into the Carhart 4-factor model. Consistent with Clarke *et al.* (2010, 2014), our results show that the inclusion of the volatility factor adds value to the factor-based performance measurement framework.[20]

3.3 Statistical tests and adjustments

The statistical significances of the differences between comparable pairs of total risk-adjusted returns are given by the p-values of the Ledoit–Wolf test, which is based on the circular block bootstrap method.[21] We also test the significance of 5-factor alphas based on their t-statistic, which is closely related to the Treynor–Black (1973) multifactor appraisal ratio, also known as the information ratio (see, e.g., Bodie *et al.*, 2010). In addition, we test the statistical significance of the differences between the decile portfolio alphas by the appropriate alpha spread test, following Kosowski *et al.* (2007). Throughout the study, we use Newey–West (1987) standard errors in the statistical tests in order to avoid problems related to autocorrelation and heteroscedasticity. In addition, we carry out Jarque and Bera's (1980) normality test for regression residuals, but the normality assumption is never violated. We also test for the existence of multicollinearity in our multifactor regression model, but the variance inflation factors do not indicate that it would be severe in any of the cases.[22]

4 The results

4.1 The results for the individual valuation ratios

The comparison between the decile portfolios formed on single selection criteria shows that the top-decile EBIT/EV portfolio generates the highest raw return (17.80% p.a.) during the 1971–2013 sample period (Table 1, Panel A). Its incremental return over the top-decile portfolios formed on the most frequently examined valuation ratios (i.e., on B/P and E/P) is remarkable, as is the corresponding incremental return of the top-decile EBITDA/EV portfolio. It is noteworthy that the three highest decile returns are all generated by EV multiples that have rarely been employed as portfolio selection criteria in previous studies. The highest value premium (9.09% based on the return difference between the extreme deciles) is given by EBITDA/EV that is exceptionally efficient in separating undervalued stocks from their overvalued counterparts. As proof of this, all the top-3 EBITDA/EV deciles significantly outperformed the market portfolio based on the Ledoit–Wolf test, whereas all the corresponding bottom-4 deciles significantly underperformed against the market portfolio (Panel B).

The best performer in terms of the Sharpe ratio is, somewhat surprisingly, the top-decile D/P portfolio, thanks to its low volatility, followed by the corresponding EBIT/EV and EBITDA/EV portfolios (Panel B). The top-decile D/P portfolio has remarkably lower volatility than the corresponding top-decile portfolios chosen on the basis of any other valuation ratio. The difference between the rankings based on raw and total risk-adjusted returns is very dramatic for the top-decile D/P portfolio, since it is the worst among the top-decile portfolios in terms of raw returns.[23] Based on the Ledoit–Wolf test statistics, the significance of outperformance against the market portfolio is the highest for the top-decile EBIT/EV portfolio

Table 1 Decile portfolio performance for 12 individual valuation ratios (May 1971–April 2013).

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *Raw returns*											
EBIT/EV	17.80	14.93	15.60	13.96	12.71	13.03	11.82	10.57	9.82	10.15	7.64
EBITDA/EV	17.76	16.43	16.33	14.77	13.05	12.69	10.80	11.05	8.85	8.66	9.09
S/EV	17.43	16.30	14.29	13.74	12.02	12.03	11.25	11.31	11.63	10.38	7.05
S/P	17.29	15.57	14.49	15.40	14.25	12.82	11.13	10.80	10.41	8.23	9.07
CF3/P	16.80	16.01	14.69	15.01	13.55	12.04	12.26	11.28	8.81	9.93	6.87
CF1/P	16.70	16.03	15.31	14.83	13.97	11.50	12.35	10.53	9.35	9.80	6.90
CFO/P	16.45	16.20	14.20	14.80	13.32	11.63	12.19	11.89	8.83	10.87	5.58
CF2/P	16.37	16.25	15.34	14.29	14.43	11.33	12.23	11.03	8.93	10.18	6.19
B/P	16.04	16.00	15.81	12.85	12.95	13.10	11.17	11.20	10.55	10.71	5.33
FCF/EV	15.29	15.49	14.04	13.40	13.52	12.82	12.38	13.01	10.44	9.98	5.31
E/P	15.17	15.75	15.08	14.29	12.49	11.57	12.93	11.28	10.31	11.52	3.65
D/P	15.04	13.93	14.16	14.81	13.36	12.44	14.33	11.81	12.42	11.09	3.95

This table presents the annualized geometric average returns (Panel A) and the Sharpe ratios (Panel B) for the decile portfolios formed on 12 valuation ratios over the period 1971–2013. Panel B also indicates the significance (in percentages in parentheses) for the performance difference between each decile portfolio and the benchmark portfolio formed on the monthly average returns of sample stocks (The significance levels are based on the Ledoit–Wolf test and italicized in case of significant underperformance). The last column shows the corresponding top–bottom-decile (D1–D10) differences, respectively (The significance levels in the last column in Panel B are for performance differences between the extreme decile portfolios). The selection criteria in each panel are in descending order according to each performance metrics of the top-decile portfolios.

Table 1 (*Continued*)

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel B: *Sharpe ratios*											
D/P	0.621	0.503	0.508	0.533	0.428	0.380	0.472	0.328	0.344	0.253	0.368
	(1.1)	(36.5)	(32.8)	(9.1)	(78.8)	(18.6)	(56.7)	*(1.2)*	(11.8)	*(0.7)*	(0.2)
EBIT/EV	0.600	0.491	0.534	0.446	0.383	0.390	0.306	0.221	0.164	0.158	0.442
	(0.2)	(8.8)	(2.8)	(49.2)	(97.7)	(91.1)	(8.9)	*(0.5)*	*(0.3)*	*(1.0)*	(0.1)
EBITDA/EV	0.597	0.561	0.573	0.484	0.394	0.370	0.256	0.248	0.128	0.107	0.490
	(0.3)	(1.3)	(0.4)	(12.5)	(92.3)	(74.2)	*(0.4)*	*(1.4)*	*(0.0)*	*(0.2)*	(0.0)
CF3/P	0.548	0.555	0.495	0.503	0.415	0.332	0.319	0.249	0.126	0.155	0.393
	(4.1)	(3.1)	(13.4)	(5.4)	(57.3)	(12.4)	(23.1)	*(3.4)*	*(0.1)*	*(0.6)*	(0.7)
CF1/P	0.542	0.558	0.526	0.495	0.436	0.306	0.323	0.219	0.146	0.149	0.393
	(5.2)	(2.5)	(6.4)	(5.9)	(39.0)	*(3.5)*	(26.8)	*(0.6)*	*(0.3)*	*(0.6)*	(0.7)
S/P	0.542	0.474	0.444	0.491	0.450	0.369	0.274	0.252	0.200	0.099	0.443
	(4.6)	(30.5)	(40.2)	(4.6)	(20.8)	(84.9)	*(1.4)*	*(2.0)*	*(2.1)*	*(0.5)*	(0.7)
CFO/P	0.534	0.569	0.480	0.500	0.406	0.314	0.316	0.265	0.130	0.190	0.344
	(4.3)	(1.8)	(21.3)	(6.5)	(90.8)	(5.1)	(25.1)	(7.8)	*(0.1)*	*(1.3)*	(0.9)
CF2/P	0.523	0.573	0.527	0.467	0.460	0.296	0.320	0.239	0.132	0.162	0.361
	(8.3)	(1.5)	(6.0)	(24.4)	(12.8)	*(3.0)*	(21.4)	*(1.9)*	*(0.1)*	*(0.9)*	(1.3)
S/EV	0.508	0.458	0.404	0.386	0.309	0.316	0.273	0.285	0.318	0.238	0.270
	(1.6)	(8.6)	(45.3)	(69.8)	(7.9)	(10.6)	*(3.2)*	(8.2)	(44.6)	(11.0)	(2.3)
FCF/EV	0.493	0.557	0.471	0.431	0.396	0.327	0.309	0.333	0.204	0.174	0.319
	(6.6)	(0.6)	(6.5)	(33.9)	(58.1)	(43.8)	(13.7)	(37.4)	*(0.0)*	*(0.0)*	(0.1)
B/P	0.484	0.522	0.524	0.374	0.368	0.359	0.269	0.259	0.226	0.202	0.282
	(18.7)	(2.2)	(0.4)	(79.5)	(59.4)	(83.1)	*(1.4)*	*(1.0)*	*(1.5)*	*(4.0)*	(6.2)
E/P	0.472	0.545	0.537	0.466	0.371	0.306	0.348	0.233	0.187	0.220	0.252
	(21.9)	(3.5)	(3.8)	(21.6)	(63.0)	(11.1)	(56.7)	*(3.2)*	*(0.6)*	*(1.7)*	(3.4)

(i.e., 0.2%). In addition, EBIT/EV is the only valuation ratio that is able to generate statistically significant difference in pairwise performance comparisons of the top-decile portfolios when it outperforms its E/P counterpart (at the 5% level).

Interestingly, the most frequently examined individual valuation ratios are not among the best criteria, but rather the reverse is true: Among the top-decile portfolios, the E/P counterpart is the worst based on the Sharpe ratio ranking, while it is the second worst in terms of raw returns. The top-decile B/P portfolio is the second worst based on Sharpe ratio and the fourth worst in terms of raw returns within the same peer group. In addition, the significances of the performance differences between the extreme decile portfolios are also lowest for these two criteria. In fact, B/P is the only criterion for which the corresponding Ledoit–Wolf test statistic is not significant at the 5% level. However, it should be noted that in terms of the Sharpe ratio, based on neither B/P nor E/P is the top-decile portfolio the best-performing decile portfolio (the same also holds for all the cash flow-based multiples). Nevertheless, our results on the relative discriminatory power of the E/P and B/P criteria are generally consistent with Dhatt *et al.* (2004) and Gray and Vogel (2012), as well as with McLean and Pontiff (2016) who state that investors learn about pricing inefficiencies from publications (According to a recent literature review of value premium and value anomalies by Pätäri and Leivo, 2017, the E/P and B/P ratios have been clearly the most frequently examined portfolio formation criteria).

4.2 The results for the combination criteria

Table 2 shows the results for the best combinations determined on the basis of the raw and total risk-adjusted returns of the top-decile portfolios (henceforth, we refer to such composite criteria as top combinations). Panel A indicates the decile returns for the 2-, 3-, and 4-combinations which generate the highest top-decile raw returns, whereas Panel B shows the results for the corresponding combinations that are the best in terms of the Sharpe ratio. Based on raw returns, both the best 2- and 3-combinations, as well as the best 4-combination can slightly add value to the selection of a value decile portfolio, compared to the individual valuation ratios. The highest decile returns (18.23%) are generated by the top-decile 2-combination portfolio formed on CF3/P and S/EV and the corresponding 3-combination portfolio formed on EBITDA/EV, S/EV, and S/P. The latter criterion is also the only one among the best combination criteria for which the decile returns

Table 2 Decile portfolio performance for the top combination criteria (May 1971–April 2013).

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *Raw returns*											
MS2 (CF3/P, S/EV)	18.23	15.09	14.74	14.90	14.14	11.68	11.95	9.82	9.80	10.03	8.20
	(0.3)	(10.0)	(8.7)	(4.0)	(24.2)	(7.6)	(12.5)	*(0.0)*	*(0.2)*	*(0.8)*	(0.1)
MS3 (EBITDA/EV,	18.23	15.33	15.19	14.96	13.56	13.38	12.69	10.77	8.19	8.08	10.15
S/EV, S/P)	(1.1)	(24.7)	(6.1)	(7.7)	(63.6)	(69.4)	(70.5)	*(0.9)*	*(0.0)*	*(0.1)*	(0.1)
MS4 (B/P, CF1/P,	18.12	16.44	15.68	13.39	13.89	11.92	11.74	11.64	8.67	8.88	9.24
EBITDA/EV, S/EV)	(0.6)	(2.1)	(2.2)	(80.2)	(52.3)	(17.1)	(6.2)	(7.5)	*(0.0)*	*(0.2)*	(0.1)
Panel B: *Sharpe ratios*											
MS4 (CFO/P, D/P,	0.629	0.514	0.503	0.411	0.328	0.325	0.263	0.275	0.250	0.203	0.426
EBITDA/EV, S/EV)	(0.6)	(22.0)	(20.6)	(64.8)	(27.3)	(32.9)	*(0.6)*	(6.3)	*(4.8)*	*(1.0)*	(0.2)
MS3 (D/P, EBITDA/EV,	0.629	0.515	0.522	0.376	0.329	0.333	0.288	0.280	0.227	0.205	0.424
S/EV)	(0.8)	(20.2)	(12.6)	(87.8)	(31.5)	(40.9)	(5.4)	(6.1)	*(1.0)*	*(1.4)*	(0.3)
MS2 (D/P, EBIT/EV)	0.617	0.483	0.523	0.357	0.377	0.308	0.287	0.277	0.236	0.240	0.377
	(2.0)	(33.1)	(8.0)	(63.5)	(85.9)	(20.4)	*(3.3)*	(7.5)	*(1.3)*	*(4.1)*	(1.0)

This table presents the annualized geometric average returns (Panel A) and the Sharpe ratios (Panel B) for the combination decile portfolios that, based on the same performance metrics, get the highest top-decile ranking within 2-, 3-, or 4-combinations. The valuation ratios included in each of these top combinations are described in the first column. Panels also indicate the significance levels (in percentages in parentheses) for the performance difference between each decile portfolio and the benchmark portfolio formed on the monthly average returns of sample stocks (The significance levels are based on the Ledoit–Wolf test and italicized in case of significant underperformance). The last column shows the corresponding top–bottom-decile (D1–D10) differences, respectively (The significance levels in the last column are for performance differences between the extreme decile portfolios). The selection criteria in each panel are in descending order according to each performance metrics of the top-decile portfolios.

decrease monotonically from the top-decile to the bottom-decile, whereas the same does not hold for any of the individual valuation ratios. Based on the total risk-adjusted top-decile performance, the best 3- and 4-combinations slightly outperform the best individual valuation ratio, whereas the best 2-combination does not. The highest overall Sharpe ratios (0.629) are documented for the 4-combination top-decile portfolio formed on CFO/P, D/P, EBITDA/EV, and S/EV and for the 3-combination top-decile portfolio formed on the three latter valuation ratios. Altogether, the EV multiples are well represented as components of the top combinations. The D/P criterion is also present in all the top combination criteria that are the best based on the total risk-adjusted performance. However, only 14.2% (i.e. 41 out of 288 feasible combinations) of the top-decile portfolios formed on the combination criteria generate higher return than the best top-decile portfolio formed on the individual valuation ratios (i.e., the top-decile EBIT/EV portfolio). Interestingly, all of these combination criteria include at least one EV multiple as a sub-criterion. In the corresponding comparisons of the total risk-adjusted returns, the similarly-calculated percentage is even lower (i.e., 0.69%), as only the two above-mentioned top-decile combination portfolios generate higher Sharpe ratio than that reported for the top-decile D/P portfolio.

Table 3 shows the average returns for all 2-, 3-, and 4-combination methods employed.[24] Interestingly, the average returns of top-decile portfolios, as well as the corresponding value premiums, decrease when moving from 2-combinations to 3-combinations, likewise when moving from 3-combinations to 4-combinations. For this particular sample, the benefits of the combination seem to be at their greatest for 2-combinations, as the added value of increasing the number of variables beyond two is negative rather than positive, on average. By contrast, Panel B indicates that there are hardly any differences in the average total risk-adjusted performance of value decile portfolios regardless of whether they are formed on 2-, 3-, or 4-combinations. However, the total risk-adjusted value premiums follow the same tendency as the value premiums based on raw returns (i.e., both decrease when moving from 2-combinations

Table 3 Average returns and Sharpe ratios for the combination portfolios.

Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *Average returns*											
MS2	16.15	15.02	14.74	13.96	13.39	12.64	11.96	11.72	10.55	10.42	5.72
MS3	15.73	14.91	14.70	13.63	13.19	12.89	11.74	12.23	10.93	10.57	5.16
MS4	15.34	14.78	14.74	13.31	13.08	12.85	11.59	12.57	11.48	10.76	4.58
Weighted average	15.64	14.88	14.72	13.55	13.18	12.83	11.72	12.28	11.10	10.62	5.02
Panel B: *Average Sharpe ratios*											
MS2	0.546	0.504	0.489	0.443	0.396	0.341	0.300	0.271	0.198	0.179	0.366
MS3	0.543	0.505	0.488	0.429	0.383	0.343	0.286	0.286	0.211	0.185	0.358
MS4	0.544	0.506	0.491	0.414	0.373	0.334	0.276	0.293	0.228	0.193	0.351
Weighted average	0.544	0.505	0.490	0.425	0.381	0.339	0.284	0.286	0.216	0.187	0.357

This table shows the annualized decile averages for the raw returns and Sharpe ratios calculated over all 2-, 3-, and 4-combinations employed (in Panels A and B, respectively). The last column shows the averages of the corresponding top–bottom-decile spreads (D1–D10). The lowest row in both panels shows the corresponding amount-weighted averages calculated over the pooled results for all three types of combinations.

to 3 combinations, and from 3-combinations to 4-combinations). Altogether, the average performance statistics shown in Table 3 are not very supportive for the combinations. For example, the average raw return of the top-decile portfolios formed on the individual valuation ratios is higher than that of the top-decile combination portfolios (16.51% vs. 15.64% p.a.). By contrast, the average top-decile Sharpe ratio is marginally higher for the combinations than it is for the individual valuation ratios (0.544 vs. 0.539). On average, the potential added value of the combinations seems to stem more from risk reduction than from return enhancement.

5 Robustness checks

5.1 The results based on the 5-factor model

As a robustness check for performance comparisons, we also calculated the abnormal decile returns for all the valuation ratios. The highest as well as the most significant 5-factor alpha (4.11% p.a. with *t*-stat 3.92) among all the decile portfolios that are based on either single selection criteria or top combinations is documented for the top-decile S/EV portfolio (Tables 4 and 5). Within the same peer group, the S/EV criterion also generates the highest as well as the most significant top–bottom-decile alpha spread (6.58% p.a. with *t*-stat 3.96). Both of these *t*-statistics are highly significant even at the 0.1% level, and S/EV is actually the only criterion that can generate the significant top–bottom alpha spread. The superior discriminatory power of the S/EV criterion in terms of alphas is further highlighted by the fact that the second-highest decile S/EV portfolio also generates highly significant (at the 1% level) positive alpha. Two other EV-based top-decile portfolios (i.e., those formed on EBIT/EV and EBITDA/EV) also generate higher alphas (3.19% and 2.95% with *t*-stats 3.78 and 3.45, respectively) than the best top combination portfolio, i.e., the top-decile portfolio formed on CF3/P and S/EV that also generated the highest raw return among all decile portfolios (the corresponding alpha is 2.93% p.a. with *t*-stat 3.49). Significantly positive top-decile alphas are also reported for the S/P criterion, as well for those 3- and 4-combination criteria that generated the highest top-decile raw returns within each type of combinations. By contrast, the alphas are not significant (at the 5% level) for any of the top-decile portfolios formed on those top combinations that generated the best top-decile total risk-adjusted performance within 2-, 3-, or 4-combinations.[25] The same also holds for the top-decile D/P portfolio which generated the highest Sharpe ratio among the decile portfolios formed on individual valuation ratios. These seemingly paradoxical differences in risk-adjusted portfolio performance underline the fact that multifactor-based alphas and the total risk-based performance metrics represent different dimensions of performance. To confirm this, we calculate the Spearman rank correlations between the Sharpe ratios and the 5-factor alpha rankings for the 18 top-decile portfolios included in Tables 4 and 5. The Spearman rho was insignificant in this case, whereas it was extremely significant between the raw return and the 5-factor alpha rankings (*t*-stat 4.69), thus indicating that at least at the ordinal level and in the top-decile comparison, the 5-factor alphas correlate more with raw returns than they do with the Sharpe ratios. By contrast, the corresponding comparison between the Sharpe ratio and raw return rankings showed that there is no consistency between these two rankings.[26]

The market excess return factor slopes range from 0.855 for the top-decile D/P portfolio to 1.109 for the D9 portfolio formed on the 2-combination of D/P and EBIT/EV, and, unsurprisingly, this factor is positive and extremely significant for all decile portfolios (Tables 1A

Table 4 5-factor decile alphas for the portfolios formed on 12 individual valuation ratios.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
S/EV	4.11	2.67	0.90	−0.10	−1.28	−1.28	−1.25	−1.65	−0.93	−2.48	6.58
	(0.0)	(0.6)	(29.6)	(88.2)	(8.3)	(8.6)	(10.0)	*(3.3)*	(29.5)	(5.6)	(0.0)
EBIT/EV	3.19	0.39	0.70	−0.99	−1.79	−1.05	−1.30	−0.33	0.50	1.44	1.75
	(0.0)	(62.1)	(31.9)	(18.7)	*(0.7)*	(14.1)	(5.2)	(68.3)	(70.1)	(37.4)	(34.0)
EBITDA/EV	2.95	1.32	0.90	−0.57	−1.30	−1.41	−1.84	−0.12	−0.02	0.95	2.00
	(0.1)	(8.8)	(18.6)	(41.4)	(7.3)	*(2.9)*	*(1.2)*	(88.0)	(98.3)	(55.4)	(27.1)
S/P	1.88	−0.15	−0.66	0.13	−0.54	−0.38	−1.53	−0.53	1.16	0.75	1.13
	(4.8)	(86.7)	(39.0)	(84.6)	(37.4)	(56.0)	*(2.4)*	(48.4)	(21.6)	(62.2)	(52.9)
CFO/P	1.19	0.61	−0.91	−0.16	−0.22	−1.51	−0.45	1.29	0.02	0.42	0.78
	(14.9)	(40.4)	(21.9)	(82.8)	(77.5)	*(2.8)*	(53.1)	(14.3)	(99.0)	(79.1)	(66.1)
B/P	1.08	0.87	1.26	−1.67	−0.93	−0.12	−1.10	−0.82	−0.40	1.17	−0.09
	(18.3)	(25.3)	(6.2)	*(1.2)*	(22.9)	(86.3)	(10.1)	(22.0)	(62.5)	(28.8)	(94.8)
CF1/P	0.94	0.26	0.26	−0.16	−0.17	−1.83	0.15	−0.65	1.17	0.54	0.39
	(26.2)	(73.4)	(73.5)	(82.0)	(80.9)	*(1.2)*	(82.4)	(45.6)	(32.8)	(73.2)	(82.6)
CF3/P	0.92	0.22	−0.11	−0.10	−0.56	−1.17	−0.16	0.25	0.57	0.54	0.38
	(26.0)	(76.7)	(88.6)	(88.1)	(40.3)	(9.7)	(81.0)	(77.4)	(63.5)	(72.8)	(82.6)
E/P	0.59	0.42	0.20	−0.23	−1.53	−1.74	0.26	1.55	0.52	0.12	0.48
	(51.1)	(57.3)	(78.4)	(74.5)	*(2.9)*	*(1.5)*	(70.4)	(12.7)	(63.4)	(93.7)	(78.1)
CF2/P	0.58	0.51	0.29	−0.64	0.17	−1.87	−0.15	−0.03	0.67	0.95	−0.37
	(48.7)	(50.7)	(70.8)	(39.9)	(79.7)	*(1.2)*	(82.4)	(96.9)	(57.1)	(54.8)	(83.6)
FCF/EV	0.30	0.49	−0.09	−0.39	0.63	0.45	0.07	0.27	−1.60	−1.29	1.59
	(66.7)	(45.0)	(88.5)	(55.3)	(38.2)	(56.6)	(92.6)	(69.5)	(8.5)	(32.2)	(28.2)
D/P	0.21	−0.93	−0.82	−0.18	−1.76	−2.66	−0.93	−2.49	−1.50	−2.06	2.26
	(78.6)	(23.4)	(36.7)	(84.6)	(5.3)	*(0.8)*	(32.6)	*(0.8)*	(13.3)	(6.0)	(9.0)

This table shows the 5-factor alphas (p.a.) for the decile portfolios formed on 12 single selection criteria (Alphas are calculated according to Eq. (1)). The selection criteria are in descending order according to the 5-factor alphas of the top-decile portfolios. The table also indicates the significance levels (in percentages in parentheses and italicized in case of significant underperformance) for both the decile alphas and the top–bottom-decile alpha spreads (in the last column).

Table 5 5-factor decile alphas for the top combination criteria.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
MS2 (CF3/P, S/EV)[a]	2.93	−0.07	−0.06	0.15	−0.02	−1.31	−0.29	−1.78	−0.46	0.17	2.76
	(0.1)	(92.2)	(92.8)	(81.1)	(97.5)	*(4.9)*	(68.3)	*(0.9)*	(60.3)	(90.8)	(10.7)
MS4 (B/P, CF1/P,	2.92	0.78	0.35	−1.29	−0.48	−0.71	−0.79	0.41	−1.28	0.14	2.78
EBITDA/EV, S/EV)[a]	(0.1)	(29.3)	(61.0)	(6.8)	(48.1)	(30.0)	(23.9)	(58.7)	(14.2)	(92.6)	(10.8)
MS3 (EBITDA/EV,	2.79	0.23	0.25	0.35	−0.42	−0.25	−0.17	−0.93	−1.74	−0.42	3.21
S/EV, S/P)[a]	(0.5)	(78.8)	(71.2)	(60.1)	(48.7)	(73.6)	(81.4)	(19.6)	*(4.6)*	(77.1)	(6.6)
MS4 (CFO/P, D/P,	1.28	−0.49	−0.75	−0.89	1.11	1.51	−0.63	0.10	−0.16	−0.97	2.25
EBITDA/EV, S/EV)[b]	(18.0)	(58.1)	(36.7)	(27.8)	(17.6)	(7.3)	(44.3)	(90.2)	(88.6)	(48.2)	(17.9)
MS3 (D/P, EBITDA/EV,	1.27	−0.37	−0.60	−1.32	1.01	1.68	0.19	−0.05	−1.01	−0.64	1.91
S/EV)[b]	(18.4)	(66.3)	(46.0)	(10.0)	(23.3)	(4.7)	(83.4)	(95.1)	(33.8)	(64.3)	(25.6)
MS2 (D/P, EBIT/EV)[b]	0.94	−0.84	−0.55	−1.66	2.03	1.19	−0.04	0.40	−1.02	−0.13	1.06
	(31.7)	(28.7)	(43.7)	(6.1)	(1.0)	(16.7)	(96.7)	(63.1)	(31.2)	(92.4)	(51.2)

This table shows the 5-factor alphas (p.a.) for the combination decile portfolios which, based on either raw returns or the Sharpe ratio, get the highest top-decile ranking within 2-, 3-, and 4-combinations (Significances in percentages are in parentheses and italicized in case of significant underperformance). Alphas are calculated according to Eq. (1). The valuation ratios included in each of these top combinations are described in the first column. The selection criteria are in descending order according to the 5-factor alphas of the top-decile portfolios. The last column shows the top–bottom-decile alpha spreads (significances in parentheses).

[a]Indicates that the criterion has generated the highest top-decile return among either 2-, 3-, or 4-combinations.

[b]Indicates the corresponding superiority in terms of the Sharpe ratio and the SKASR.

and 2A in Appendix). Expectedly, the next highest proportion of significant factor exposures is documented for the HML factor which is significant for 154 out of 180 portfolios (as shown in Panel B of Table 1A, it is significantly positive in 83 cases and significantly negative in 71 cases). The HML slope is positive and highly significant for all the top-decile portfolios (Panel A of Table 2A), whereas it is significantly negative (at the 5% level) for all the bottom-decile portfolios other than those formed on the S/EV and FCF/EV, for both of which HML is only weakly significant (Panel B of Table 2A).

In terms of occurrence frequencies of significant factor exposures, HML is followed by the volatility factor (SMV) that is significant for 140 out of 180 portfolios (Panel B of Table 1A shows that it is significantly positive (negative) in 91 (49) cases). Among the top-decile portfolios, it is positive in 17 out of 18 cases, of which 10 are also significant (at the 5% level), indicating that these portfolios are tilted towards low-volatility stocks. In line with the subsequent results of this study, the most significant low-volatility tilt is documented for the top-decile D/P portfolio. By contrast, and also parallel to the rest of the results, the only high-volatility tilt among the top-decile portfolios is documented for S/EV, for which it is also highly significant (at the 0.01% level). The top-decile S/EV portfolio is very exceptional in this sense, as among the 72 top-4 decile portfolios, there are only two portfolios with significantly negative SMV slopes. Instead, all the other bottom-decile portfolios, except those formed on D/P and S/EV, have significant high-volatility exposures. These two bottom-decile portfolios are also exceptional in the sense that their SMV slopes are positive, though insignificant. At more general level, the low-volatility exposures are concentrated on the six highest decile portfolios (in 85 out of 91 significantly positive cases), whereas the significant tilts towards the high-volatility stocks are mostly documented for the three lowest decile portfolios (in 38 out of 49 cases).

The addition of the volatility factor into the Carhart 4-factor model has also interesting impacts on the other factors, and particularly, on the momentum factor. Although based on the Carhart 4-factor model, the momentum factor is significant for 54 out of 180 portfolios, it is significant in 101 out of 180 cases after the inclusion of the volatility factor! Moreover, the extension also increases the number of significant SMB slopes from 77 to 79, while marginally reducing the proportion of significant HML by 6 to 154, thereby indicating that the volatility factor is clearly a distinct factor that hardly erodes any explanatory power of other four factors.

Of 101 significant momentum slopes, 42 are positive, whereas 59 are negative (Panel B of Table 1A). As value and momentum are negatively correlated (see, e.g., Asness *et al.*, 2013), it is not surprising that significantly negative momentum exposures are concentrated at the high end of deciles, whereas positive momentum exposures are typical for the deciles belonging to the lower half of the sample. All the top-decile momentum slopes are negative, and also significant in 14 out of 18 cases (The exceptions are the top-decile portfolios formed on S/EV, and those formed on the top combinations that generated the best top-decile total risk-adjusted performance within either 2-, 3- or 4-combinations). By contrast, all the corresponding bottom-decile slopes, except the WML of FCF/EV that is insignificant, are positive (38 out of 42 positive momentum exposures are within the four lowest deciles, whereas 53 out of 59 significantly negative WML slopes are documented for the top-4 decile portfolios). As the momentum factor is significant for over half of the decile portfolios, whether it is positive or negative, it is important to be included in the factor model used for performance evaluation of the decile portfolios even though the momentum is not included in the portfolio-formation criteria.

Of all the factors included in the performance measurement framework, the significant slopes of size factors are most evenly dispersed over the deciles. However, significantly positive and negative SMB slopes differ in that the majority of significantly positive SMB exposures are documented for either the top-2 or bottom-2 decile portfolios, whereas the majority of negative size exposures are documented for the rest six intervening decile portfolios. Among the top-decile portfolios, the most significant positive tilt towards small-cap stocks is documented for the corresponding EBIT/EV portfolio. Also, the top-decile portfolios formed on all the other EV multiples and on B/P have significantly positive SMB slopes. By contrast, reverse tilts towards large-caps are reported only for those 2-, 3-, and 4-combinations that are the top combinations in terms of total risk-adjusted returns. As among the upper-decile portfolios formed on individual valuation ratios, the corresponding D/P portfolios have the strongest larger-cap tilts, and in addition, as D/P is also included in all the just-mentioned top combinations as one subcriterion, its inclusion shifts the aggregate SMB exposure of the constituent stocks towards larger-cap stocks to the extent that the negative SMB exposure for such combinations is even stronger (and more significant) than it is for the upper-decile D/P portfolios. Somewhat surprisingly, occurrence frequency for significant factor exposures is the lowest for the SMB, although it is a significant explanatory factor for 79 out of 180 portfolios (Panel B of Table 1A shows that it is significantly positive (negative) in 36 (43) cases).

5.2 *The sub-period results*

As the next robustness check, we examine whether the top-decile rankings of valuation ratios have any consistency over time by dividing the full-length sample period into two sub-periods of equal length. The test reveals the persistence in return rankings (see Table 6). During both 21-year sub-periods, the three EV-based multiples (EBIT/EV, EBITDA/EV, and S/EV) generated higher returns than any of the price-based multiples except S/P, which, during the latter sub-period, is the second best in terms of raw returns and generates the highest top–bottom-decile return spread. Although the highest decile returns are clearly lower during the latter 21-year sub-period, the average value premiums and the incremental returns (over the sample average return) of the best decile portfolios do not decrease to the same extent. In fact, the average excess returns (over risk-free rate) of the top-decile portfolios increase by 2.96% p.a. during the latter sub-period as a result of lower interest rates. In addition to this, as the average total risk of the top-decile portfolios is somewhat lower during the latter sub-period, the Sharpe ratios (not reported here) are higher during that period for all 12 top-decile portfolios. The same also holds for the SKASRs even though the distribution of the market returns is slightly more negatively skewed during the latter 21-year period.[27] In light of these results, the value anomalies seem to have been persistent in the long term.

The *t*-statistic of the Spearman rank correlation coefficient for the top-decile return rankings is 5.89, which is a highly significant (even at 0.02% level) indication of the persistence in return rankings over the 21-year sub-periods. The corresponding *t*-statistic for the Sharpe ratio (SKASR) rankings is 2.47 (2.30), which both are significant at the 5% level and show that the performance persistence also exists based on the total risk-adjusted performance rankings (Panel A, Table 7). To test the shorter-term performance persistence, we further divide the 42-year sample period into six 7-year sub-periods and fourteen

Table 6 Decile returns for 12 individual valuation ratios over 21-year sub-periods.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *raw returns* (*May* 1971–*April* 1992)											
EBIT/EV	19.27	16.42	16.28	14.58	12.98	14.29	12.72	10.17	10.05	12.10	7.18
S/EV	18.62	16.73	14.84	14.17	14.30	13.40	12.06	11.23	11.78	11.80	6.82
EBITDA/EV	18.58	17.15	17.93	16.31	12.23	13.82	10.54	12.25	9.75	10.36	8.22
S/P	17.75	17.82	15.14	16.02	14.29	13.42	10.83	11.47	11.18	10.99	6.76
CF1/P	17.70	17.38	16.58	16.66	14.18	11.24	13.49	11.39	10.52	9.67	8.03
CFO/P	17.46	18.80	15.26	16.09	13.20	10.80	14.00	13.11	10.06	10.07	7.39
CF3/P	17.45	17.70	15.78	17.18	13.67	11.89	13.85	11.91	9.88	9.54	7.92
CF2/P	17.14	17.73	16.72	15.42	15.37	11.16	13.26	11.98	9.50	10.47	6.67
B/P	17.05	17.48	18.02	13.63	13.13	13.73	9.83	11.42	11.96	12.58	4.47
D/P	16.95	15.93	16.25	17.22	15.74	12.71	15.14	13.24	12.28	11.04	5.90
E/P	16.53	17.70	16.75	15.44	13.00	11.69	13.45	13.06	10.63	10.69	5.84
FCF/EV	15.70	17.59	14.98	14.28	13.77	13.69	13.18	12.59	10.54	12.53	3.16
Panel B: *Raw returns* (*May* 1992–*April* 2013)											
EBITDA/EV	16.94	15.70	14.75	13.25	13.88	11.57	11.07	9.86	7.95	7.00	9.95
S/P	16.82	13.35	13.83	14.77	14.20	12.21	11.42	10.11	9.64	5.52	11.30
EBIT/EV	16.34	13.45	14.93	13.35	12.44	11.79	10.92	10.96	9.58	8.24	8.10
S/EV	16.24	15.86	13.74	13.31	9.79	10.68	10.44	11.41	11.48	8.98	7.26
CF3/P	16.14	14.32	13.61	12.86	13.41	12.18	10.68	10.63	7.73	10.32	5.82
CF1/P	15.69	14.69	14.05	13.03	13.76	11.75	11.21	9.68	8.18	9.91	5.78
CF2/P	15.59	14.77	13.96	13.14	13.47	11.48	11.19	10.07	8.34	9.87	5.72
CFO/P	15.44	13.64	13.15	13.51	13.43	12.45	10.39	10.67	7.60	11.67	3.78
B/P	15.01	14.52	13.61	12.06	12.76	12.45	12.52	10.97	9.15	8.85	6.16
FCF/EV	14.88	13.41	13.10	12.53	13.27	11.95	11.57	13.43	10.35	7.48	7.39
E/P	13.82	13.83	13.43	13.14	11.98	11.45	12.40	9.54	9.99	12.35	1.47
D/P	13.16	11.96	12.11	12.45	11.04	12.17	13.53	10.40	12.55	11.14	2.02

Panel A presents the annualized geometric average returns for the decile portfolios formed on 12 valuation ratios over the period May 1971–April 1992, while Panel B does the same for the period May 1992–April 2013. The selection criteria in each panel are in descending order according to top-decile returns. The last column shows the top–bottom-decile return spreads (D1–D10).

3-year sub-periods and assess the consistency of the sub-period raw return rankings based on Kendall's W test (Panels B and C, Table 7). The coefficient of concordance for the 7-year periods is significant at the 1% level and weakly significant for the 3-year periods (significance level is 9.96%). As expected, the persistence of the sub-period rankings deteriorates when the length of the sub-periods is shortened.

To find out whether the differences in the sensitivity to rising or declining stock market returns explain the outperformance (underperformance) of the best (worst) portfolios, we divided the full sample period into bullish and bearish months based on the sign of the market return calculated from the first Fama–French (1993) factor (i.e., excess market return), in line with Fuller and Goldstein (2011). Based on that criterion, the full

Table 7 Sub-period rankings for the top-decile portfolios formed on 12 valuation ratios.

Sub-period	B/P	CF1/P	CF2/P	CF3/P	CFO/P	D/P	E/P	EBIT/EV	EBITDA/EV	FCF/EV	S/EV	S/P
Panel A: Sharpe ratios rankings over the 21-year sub-periods												
5/1971–4/1992	10	6	9	8	7	1	11	2	3	12	4	5
5/1992–4/2013	10	5	6	4	7	1	11	3	2	9	12	8
Panel B: Raw return rankings over the 7-year sub-periods												
5/1971–4/1978	4	5	7	6	8	3	9	2	1	11	10	12
5/1978–4/1985	12	6	8	4	7	10	9	3	5	11	2	1
5/1985–4/1992	11	6	7	9	5	12	8	2	4	10	1	3
5/1992–4/1999	7	8	10	3	5	12	11	4	2	1	6	9
5/1999–4/2006	9	7	4	6	8	12	11	5	3	10	1	2
5/2006–4/2013	9	8	10	7	5	2	6	4	3	12	11	1
Panel C: Raw return rankings over the 3-year sub-periods												
5/1971–4/1974	3	2	6	7	5	1	9	8	4	12	11	10
5/1974–4/1977	3	10	11	6	5	9	4	1	2	8	7	12
5/1977–4/1980	11	2	4	1	7	12	9	8	3	10	5	6
5/1980–4/1983	9	5	6	4	10	11	12	3	7	8	1	2
5/1983–4/1986	12	8	9	10	3	6	2	1	5	11	7	4
5/1986–4/1989	2	1	3	5	9	12	11	8	7	10	4	6
5/1989–4/1992	10	8	12	11	5	1	7	4	6	2	3	9
5/1992–4/1995	5	4	8	7	6	12	11	2	1	3	9	10
5/1995–4/1998	10	8	9	4	5	12	7	6	3	1	11	2
5/1998–4/2001	3	10	8	7	11	4	12	6	5	2	1	9
5/2001–4/2004	10	4	3	6	7	12	9	5	8	11	1	2
5/2004–4/2007	10	3	4	2	5	12	8	7	6	11	9	1
5/2007–4/2010	9	6	7	5	2	8	3	1	4	11	12	10
5/2010–4/2013	4	6	7	8	10	2	11	9	5	12	3	1

Panel A shows the top-decile Sharpe ratio rankings for 12 valuation ratios over two subsequent 21-year sub-periods, whereas Panel B (C) indicates the raw return rankings for each of the 7-year (3-year) sub-periods.

42-year sample period includes 308 bullish and 196 bearish months. Panel A of Table 8 shows the average bearish-month excess returns above or below the corresponding sample average (that is −4.11% p.m.) for the decile portfolios formed on single selection criteria, among which the top-decile D/P portfolio is clearly superior to all the others.[28] The superiority of the same portfolio also holds over the six top combinations included in Panel B, although the return differences in favor of the top-decile D/P portfolio are remarkably smaller in those cases. Among the top combination criteria, the smallest average losses during the bearish months are reported for the top-decile portfolios formed on those criteria in which D/P is included. By contrast, during the bullish months, for which the sample average return is 4.44% p.m., the top-decile D/P portfolio generates the lowest average return among the top-decile portfolios (Panels A and B in Table 9). The return rank order of single selection criteria vary between the bullish and bearish months to the extent that the

Table 8 Decile portfolio returns for the single and the top combination criteria in bearish months.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *Monthly raw returns for the single selection criteria*											
D/P	1.88	1.28	1.40	1.15	0.78	0.77	0.77	0.31	0.12	−0.51	2.38
EBIT/EV	0.87	0.73	0.93	0.61	0.45	0.33	−0.19	−0.84	−1.51	−1.84	2.71
EBITDA/EV	0.78	0.87	0.97	0.66	0.53	0.24	−0.30	−0.63	−1.64	−1.99	2.77
CFO/P	0.74	1.01	0.91	0.80	0.35	0.10	−0.25	−0.92	−1.57	−1.60	2.34
CF3/P	0.72	0.98	0.83	0.77	0.40	0.14	−0.20	−0.76	−1.63	−1.71	2.43
CF1/P	0.71	0.98	0.91	0.77	0.44	0.09	−0.21	−0.79	−1.61	−1.76	2.47
CF2/P	0.68	0.99	0.88	0.78	0.47	0.05	−0.19	−0.77	−1.60	−1.75	2.43
E/P	0.60	0.98	1.11	0.66	0.34	0.01	−0.23	−1.09	−1.33	−1.47	2.07
FCF/EV	0.59	0.96	0.82	0.57	0.09	−0.34	−0.42	−0.38	−0.94	−1.34	1.93
S/P	0.55	0.49	0.45	0.49	0.51	0.25	−0.20	−0.36	−1.00	−1.60	2.15
B/P	0.49	0.66	0.75	0.33	0.22	−0.04	−0.27	−0.56	−0.72	−1.22	1.70
S/EV	−0.04	−0.07	0.02	0.06	−0.10	0.02	−0.20	−0.10	0.12	−0.10	0.06
Panel B: *Monthly raw returns for the top combination criteria*											
MS2 (D/P, EBIT/EV)[b]	1.75	1.11	1.10	0.45	−0.10	−0.58	−0.69	−0.91	−1.20	−1.30	3.05
MS3 (D/P, EBITDA/EV, S/EV)[b]	1.58	1.15	1.02	0.56	−0.14	−0.57	−0.59	−0.98	−1.09	−1.31	2.89
MS4 (CFO/P, D/P, EBITDA/EV, S/EV)[b]	1.56	1.13	0.99	0.66	−0.20	−0.57	−0.70	−0.92	−1.04	−1.28	2.84
MS4 (B/P, CF1/P, EBITDA/EV, S/EV)[a]	0.78	0.82	0.90	0.51	0.37	−0.02	−0.31	−0.60	−1.20	−1.70	2.48
MS2 (CF3/P, S/EV)[a]	0.68	0.61	0.61	0.51	0.27	0.02	−0.26	−0.64	−0.77	−1.48	2.16
MS3 (EBITDA/EV, S/EV, S/P)[a]	0.65	0.58	0.49	0.40	0.29	0.17	−0.05	−0.59	−0.89	−1.49	2.14

This table presents the average bearish-month excess returns (above (+) or below (–) the sample average) for the decile portfolios formed on 12 valuation ratios (Panel A) and the top combination criteria that generate either the highest raw returns or the highest Sharpe ratio within 2-, 3-, and 4-combinations, respectively (Panel B). The corresponding top–bottom-decile return spreads are reported in the last column. The selection criteria in each panel are in descending order according to the average returns of the top-decile portfolios.
[a]Indicates that the criterion has generated the highest top-decile return among 2-, 3-, and 4-combinations.
[b]Indicates the corresponding superiority in terms of the Sharpe ratio and the SKASR.

Spearman rank correlation coefficient is negative (t-stat −1.41), although insignificant. This finding supports the idea of seeking benefits from combining individual selection criteria.

The average bullish-month returns for the top combinations are shown in Panel B in Table 9. In bullish months, the highest average return among the top-decile portfolios is reported for the 3-combination of EBITDA/EV, S/EV, and S/P, whereas the same portfolio is on the bottom in the bearish-month return ranking. These results are not surprising as its constituents generate the top-3 bullish-month returns, and based on bearish-month average returns, two of them are among the worst three in the corresponding ranking

Table 9 Decile portfolio returns for the single and the top combination criteria in bullish months.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: Monthly raw returns for the single selection criteria											
S/EV	0.57	0.46	0.15	0.04	−0.06	−0.14	−0.09	−0.15	−0.26	−0.27	0.84
S/P	0.15	−0.02	−0.13	−0.04	−0.20	−0.20	−0.10	−0.03	0.36	0.50	−0.35
EBITDA/EV	0.04	−0.18	−0.26	−0.24	−0.36	−0.21	−0.08	0.18	0.61	0.84	−0.79
B/P	0.04	−0.09	−0.17	−0.25	−0.16	0.03	−0.05	0.15	0.18	0.55	−0.52
EBIT/EV	−0.01	−0.27	−0.32	−0.30	−0.35	−0.23	−0.02	0.27	0.65	0.93	−0.94
CF3/P	−0.03	−0.31	−0.37	−0.29	−0.21	−0.22	0.04	0.31	0.60	0.80	−0.83
CF1/P	−0.04	−0.30	−0.34	−0.31	−0.19	−0.26	0.06	0.23	0.65	0.82	−0.86
CF2/P	−0.05	−0.28	−0.32	−0.38	−0.15	−0.25	0.03	0.28	0.60	0.86	−0.92
CFO/P	−0.08	−0.30	−0.48	−0.33	−0.21	−0.25	0.07	0.50	0.56	0.85	−0.93
FCF/EV	−0.12	−0.35	−0.44	−0.35	0.00	0.21	0.21	0.26	0.32	0.55	−0.67
E/P	−0.14	−0.34	−0.51	−0.30	−0.30	−0.19	0.15	0.54	0.58	0.84	−0.98
D/P	−1.03	−0.76	−0.82	−0.57	−0.49	−0.60	−0.37	−0.37	−0.16	0.11	−1.14
Panel B: Monthly raw returns for the top combination criteria											
MS3 (EBITDA/EV, S/EV, S/P)[a]	0.19	−0.11	−0.07	−0.04	−0.13	−0.07	−0.01	0.13	0.00	0.41	−0.21
MS2 (CF3/P, S/EV)[a]	0.17	−0.17	−0.21	−0.12	−0.05	−0.18	0.04	0.03	0.12	0.65	−0.48
MS4 (B/P, CF1/P, EBITDA/EV, S/EV)[a]	0.09	−0.14	−0.29	−0.30	−0.15	−0.13	0.05	0.24	0.27	0.66	−0.57
MS4 (CFO/P, D/P, EBITDA/EV, S/EV)[b]	−0.76	−0.58	−0.45	−0.42	0.06	0.40	0.31	0.55	0.58	0.62	−1.38
MS3 (D/P, EBITDA/EV, S/EV)[b]	−0.79	−0.60	−0.43	−0.43	0.03	0.42	0.30	0.62	0.53	0.65	−1.44
MS2 (D/P, EBIT/EV)[b]	−0.97	−0.66	−0.52	−0.39	0.12	0.35	0.38	0.55	0.66	0.79	−1.76

This table presents the average bullish-month excess returns (above (+) or below (–) the sample average) for the decile portfolios formed on 12 valuation ratios (Panel A) and the top combination criteria that generate either the highest raw returns or the highest Sharpe ratio within 2-, 3-, and 4-combinations, respectively (Panel B). The corresponding top–bottom-decile return spreads are reported in the last column. The selection criteria in each panel are in descending order according to the average returns of the top-decile portfolios.
[a]Indicates that the criterion has generated the highest top-decile return among 2-, 3-, and 4-combinations.
[b]Indicates the corresponding superiority in terms of the Sharpe ratio and the SKASR.

comparisons of individual valuation ratios. In contrast to D/P, S/EV seems to represent the opposite end among the top-decile portfolios as it clearly generates the highest average bullish-month return, as well as the lowest average bearish-month return (this conjecture was also confirmed based on the un-tabulated results of pairwise correlation comparisons of individual valuation ratios: the lowest return correlation was documented between D/P and S/EV). In this sense, the top-decile S/EV portfolio has the most momentum-type characteristics among the top-decile portfolios and it also has the strongest and highly significant return correlation with pure momentum decile portfolios available in French's data library.[29] The S/EV criterion is also exceptional in that whereas the outperformance of the value portfolios formed on other valuation ratios over the market portfolio is attributed to their outperformance during the bearish months, the outperformance of the S/EV-based value portfolios is based on their outperformance during the bullish months.

Although the 42-year sample period is dominated by bullish months (308 out of 504 months), the more decisive issue for the full sample-period total risk-adjusted performance of decile portfolios is how they perform during the bearish months. As evidence of this, the order of the single selection-criteria ranking formed on the basis of the average bearish-month returns of top-decile portfolios is almost the same as that formed on the Sharpe ratio (as well as on the SKASR) for the full sample period. In addition, the return differences between the extreme deciles are generally much larger during the bearish months than during the bullish months. Moreover, on average, value stocks tend to outperform glamour stocks during the bearish months more than glamour stocks outperform value stocks during the bullish months. Again, S/EV is exceptional, as during the bullish months, the top-decile S/EV portfolio generates higher returns than its bottom-decile counterpart. More generally, our results are in line with Lakonishok *et al.* (1994), Bird and Whitaker (2003), and Hwang and Rubesam (2013), and Pätäri *et al.* (2016), who all conclude that value stocks outperform glamour stocks particularly during bad times.[30]

5.3 *Decile portfolio performance for the non-exclusive samples*

Finally, we test the sensitivity of the rank order of individual valuation ratios to the requirement

Table 10 Decile portfolio performance for the non-exclusive samples.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: *Raw returns*											
EBITDA/EV	17.99	17.14	16.50	13.88	12.87	11.94	10.75	9.35	7.18	7.84	10.15
S/EV	17.59	16.18	14.37	13.72	11.27	11.29	11.53	11.52	11.76	10.36	7.22
EBIT/EV	17.47	15.52	15.41	13.55	12.47	12.21	10.70	10.35	8.05	9.44	8.03
S/P	17.30	16.01	14.43	15.08	14.04	12.56	11.52	10.46	9.09	7.91	9.39
CF1/P	17.04	16.42	14.52	15.22	13.86	11.14	11.37	9.62	9.75	8.60	8.45
CF3/P	16.94	16.01	15.24	14.33	13.30	11.79	11.41	9.78	8.43	9.61	7.33
CF2/P	16.74	16.22	14.95	14.51	14.00	10.97	11.26	9.65	9.16	9.27	7.48
CFO/P	16.26	16.14	14.36	14.84	12.92	12.00	11.18	9.92	8.67	10.55	5.71
B/P	16.10	15.93	14.55	13.61	12.70	12.76	11.13	10.75	9.32	10.85	5.25
FCF/EV	15.94	15.76	13.90	13.75	12.30	12.55	12.37	12.80	9.80	9.52	6.42
E/P	15.44	15.48	14.49	14.65	12.39	12.20	11.03	9.95	9.80	11.55	3.89
D/P	14.22	14.32	14.35	14.30	14.13	12.76	13.50	12.16	12.52	11.04	3.18

This table presents the annualized geometric average returns (Panel A) and the Sharpe ratios (Panel B) for the decile portfolios formed on 12 single selection criteria over the period May 1971–April 2013 without the restrictive condition that, to be included in the sample, a firm must have all 12 valuation ratios calculable in each portfolio-formation time point. Panel B also indicates the significance levels (in percentages in parentheses) for the performance differences between each decile portfolio and the benchmark portfolio formed on the monthly average returns of sample stocks (The significance levels are based on the Ledoit–Wolf test and italicized in case of significant underperformance). The last column in Panel A shows the top–bottom-decile return spreads (D1–D10), whereas the last column in Panel B indicates the performance differences between the extreme decile portfolios with the significance levels. The selection criteria in each panel are in descending order according to each performance metrics of the top-decile portfolios.

Table 10 (*Continued*)

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel B: *Sharpe ratios*											
D/P	0.618	0.565	0.555	0.511	0.482	0.397	0.429	0.347	0.344	0.254	0.364
	(5.5)	(12.0)	(21.4)	(31.9)	(70.0)	(17.8)	(67.4)	*(1.5)*	(5.4)	*(0.6)*	(0.7)
EBITDA/EV	0.608	0.600	0.588	0.457	0.390	0.332	0.257	0.166	0.066	0.081	0.527
	(0.0)	(0.2)	(0.2)	(17.4)	(84.6)	(49.6)	*(0.9)*	*(0.1)*	*(0.0)*	*(0.1)*	(0.0)
EBIT/EV	0.585	0.521	0.524	0.432	0.368	0.344	0.243	0.198	0.098	0.134	0.451
	(0.1)	(1.6)	(1.8)	(31.3)	(91.3)	(71.8)	*(1.2)*	*(1.3)*	*(0.0)*	*(0.6)*	(0.0)
CF1/P	0.565	0.591	0.487	0.516	0.434	0.283	0.271	0.167	0.155	0.109	0.456
	(2.4)	(0.7)	(14.8)	(1.8)	(30.9)	*(2.2)*	*(4.3)*	*(0.2)*	*(0.8)*	*(0.1)*	(0.1)
CF3/P	0.556	0.568	0.526	0.473	0.405	0.315	0.273	0.172	0.108	0.144	0.412
	(2.8)	(1.3)	(3.0)	(9.8)	(48.7)	(13.5)	(5.4)	*(0.3)*	*(0.1)*	*(0.4)*	(0.4)
CF2/P	0.544	0.583	0.512	0.480	0.439	0.276	0.270	0.167	0.134	0.131	0.413
	(3.2)	(0.9)	(4.8)	(8.2)	(17.0)	*(1.5)*	*(3.9)*	*(0.2)*	*(0.4)*	*(0.3)*	(0.5)
FCF/EV	0.534	0.585	0.476	0.445	0.331	0.313	0.313	0.325	0.180	0.157	0.377
	(1.3)	(0.1)	(5.6)	(13.9)	(39.4)	(33.8)	(16.3)	(29.8)	*(0.0)*	*(0.0)*	(0.0)
S/P	0.530	0.505	0.450	0.490	0.454	0.364	0.306	0.239	0.151	0.087	0.443
	(8.0)	(11.2)	(27.6)	(5.0)	(15.5)	(69.3)	(11.6)	*(1.9)*	*(0.1)*	*(0.3)*	(0.9)
CFO/P	0.525	0.580	0.495	0.502	0.389	0.325	0.257	0.172	0.121	0.175	0.350
	(4.5)	(0.5)	(10.5)	(1.9)	(87.4)	(27.8)	*(2.5)*	*(0.3)*	*(0.1)*	*(0.8)*	(0.7)
B/P	0.502	0.556	0.487	0.426	0.367	0.345	0.265	0.239	0.170	0.199	0.303
	(11.1)	(0.6)	(3.1)	(29.0)	(73.8)	(59.4)	*(0.7)*	*(0.3)*	*(0.1)*	*(4.6)*	(4.9)
S/EV	0.499	0.448	0.400	0.386	0.278	0.282	0.289	0.300	0.349	0.247	0.252
	(2.7)	(7.9)	(37.9)	(80.2)	*(0.8)*	*(1.3)*	(7.5)	(14.6)	(74.8)	(13.5)	(3.3)
E/P	0.489	0.540	0.508	0.484	0.365	0.328	0.243	0.174	0.162	0.215	0.274
	(11.8)	(2.0)	(5.5)	(6.2)	(87.1)	(48.6)	*(1.7)*	*(0.5)*	*(0.5)*	*(2.5)*	(2.5)

that, in order to be included in the decile division for the subsequent 1-year holding period, a firm must have all 12 valuation ratios available in each portfolio-formation checkpoint. Table 10 shows that when this requirement is removed, the results do not change much. The average return of top-decile portfolios remains practically the same, although the average sample return calculated over 12 different samples decreases somewhat (by 31 basis points) compared to the average sample return when the exclusive requirement holds. This implies that the average return of the stocks that fulfill the requirement is higher than that of the stocks that do not fulfill it. However, the average value premium is larger for the sample with no such restriction (7.21% vs. 6.39% p.a.), which indicates that the results for our main sample are robust to this kind of sample selection bias. The Spearman rank correlation test also shows that the top-decile performance rankings based on our exclusive sample and on non-exclusive samples are nearly identical (the corresponding test statistics are 11.64 (for raw return rankings) and 6.35 (for the Sharpe ratio rankings), which are both extremely significant).

6 Discussion

Overall evidence indicates that there are remarkable differences in the characteristics between the valuation ratios, and some criteria are superior to others in selecting the long-only portfolios, as well as for the basis of long–short strategies. We find evidence of the strong relative efficacy of three EV multiples (EBIT/EV, EBITDA/EV, and S/EV), indicating that these valuation ratios used commonly by corporate acquirers are also useful for the value investors. Similar results for the two first-mentioned EV multiples were also reported by Gray and Vogel (2012), although the spectrum of valuation ratios examined was much narrower in their study than it is in ours. By contrast, the S/EV criterion has not been examined at all in previous financial literature as a stand-alone criterion for this purpose, to the best of our knowledge. S/EV is also the most unique among the valuation ratios as the S/EV-based return difference factor does not correlate with either the HML or the WML factors. Among the value portfolios, their S/EV-based counterparts also have the strongest return correlation with the corresponding momentum portfolios, as well as some other momentum-type characteristics, such as return-generation pattern during the bearish and bullish months. Therefore, S/EV is a good supplementary valuation criterion for a value investor who is not willing to combine value criteria with momentum indicators, but would like to have momentum-type characteristics embedded in the composite value criteria. In addition, the top-decile S/EV portfolio generates the highest and the most significant overall 5-factor alpha. Moreover, the highest and the most significant top–bottom-decile alpha spread is documented for the S/EV criterion.

By contrast, the top-decile D/P portfolios represent the other extreme in the sense that their volatilities, as well as downside risks, measured as average losses during the bearish months as well as during the bearish periods, are remarkably lower than those of any of the other decile portfolios examined.[31] Although the D/P value portfolios are among the worst rather than among the best in raw return-based top-decile rankings, their relative performance looks totally different in terms of total risk-adjusted returns. The high dividend yield seems to offer some protection against declining stock prices. The low-risk characteristic of the top-decile D/P portfolio also extends to the combination criteria in which D/P is included; in each of the 2-, 3-, and 4-combinations that are the best in terms of total risk-adjusted returns, D/P is among the sub-criteria. One distinctive feature of the high-D/P portfolios is their strong tilt towards larger-cap stocks,[32] whereas other higher decile portfolios formed on other valuation ratios have either a significant small-cap tilt (i.e., the top-decile B/P, FCF/EV, EBIT/EV, EBITDA/EV, and S/EV portfolios) or insignificant SMB exposure. The tendency of D/P to prefer larger-cap stocks in the value portfolios is so pervasive that the best combination criteria including D/P as one sub-criterion are even more significantly tilted towards larger-cap stocks than the pure D/P value portfolio, even if the other sub-criteria have significant small-cap tilts.

The results show that none of the cash flow multiples is superior to another, since their rank orders based on either top-decile portfolio performance or the corresponding value premium depend on the performance metrics employed.

In comparisons between the top-decile portfolios formed on cash flow multiples and E/P, the latter is dominated by all the former-type multiples in all pairwise comparisons with the exception that the top-decile CF2/P and FCF/EV portfolios generate a slightly lower, though insignificant, alpha than their E/P-based counterpart.

The constituent stocks included in each decile portfolio were equal-weighted at annual portfolio-formation points, but weight changes stemming from return differences during 1-year holding periods were taken into account in the calculation of monthly time-series returns of decile portfolios. This methodology was followed instead of value-weighting the returns because the former is more realistic from the viewpoint of practical portfolio management, which is the focus of this study (see Endnote 13 for the motivation for the chosen methodology). Based on the recent literature (e.g., see Loughran and Wellman, 2011; Gray and Vogel, 2012) and our findings of significant divergence of market-cap tilts among valuation ratios, the results for the value-weighted returns would most likely have differed from ours.

Overall evidence of the benefits of using combination criteria is mixed with the respect that in terms of raw and total risk-adjusted returns, the best combination criteria have added value to portfolio selection, compared to all individual valuation criteria examined. By contrast, based on the 5-factor alphas, the added value of the same combination criteria is somewhat restricted, extending only to the price-based valuation ratios, but not to the best EV-based multiples. In addition, it should be noted that altogether, we formed 288 combinations, among which we chose six to represent the top combinations (i.e., those that generate either the highest raw return or the Sharpe ratio within each of the 2-, 3- and 4-combinations). However, it is very unlikely that the investor would have been able to select these particular combinations *ex ante* among all 288 combinations being examined.[33] Moreover, the average performance statistics of the top-decile portfolios calculated over all feasible combinations shows that their average performance is only mediocre at its best. Therefore, the added value of forming combinations within the value dimension is limited and, in many cases, non-existent at least for this particular sample data, and based on this particular combination method. With this respect, our results are in line with Bird and Casavecchia (2007) who find no evidence of added value from combining valuation ratios in the European stock markets.

In addition to valuation ratios, other selection criteria have also been shown to enhance value portfolio performance. First and foremost, abundant evidence for the benefits of combining value and momentum has been presented (see, e.g., Asness *et al.*, 2013; Fisher *et al.*, 2016). In addition, the added value of including profitability measures in value portfolio selection is documented by Novy-Marx (2013) and Fong and Ong (2016), whereas Bartov and Kim (2004) and Simlai (2016) find that the combination of high B/P and low accruals would have boosted the annual return in comparison to the returns from the corresponding single selection criteria. However, the focus of our analysis was on the comparison and combination of individual valuation ratios; therefore, these kinds of extensions were beyond the scope of this study, although they are an interesting subject for further research. Nevertheless, our results showed that it is not irrelevant what valuation ratio(s) is (are) used for value portfolio selection. Particularly, D/P proved to be superior for risk reduction purposes, whereas the three EV multiples (EBIT/EV, EBITDA/EV, and S/EV) generated, in most cases, somewhat higher raw returns, as well as higher risk-adjusted top-decile returns, than the price-based multiples. The same three EV multiples were also well-represented in the best combination criteria. With respect to the combination methods employed, we concentrate on one simple variation of the combination methods. Hence, the results of this study do not exclude the possibility that the benefits of combination strategies could have been higher had we used either more sophisticated or simpler combination methods.[34]

Like most 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 exclusion causes a small upward bias in the performance metrics of decile portfolios, recent evidence shows that such a bias is marginal for low-turnover strategies like those examined in this study (see, e.g., Frazzini *et al.*, 2015; Novy-Marx and Velikov, 2016), as all our decile portfolios are updated and rebalanced once a year only. In addition, many of the stocks in a certain decile continue to belong to the same decile after the reformation of the portfolios, and in such cases, only rebalancing trades instead of sale or purchase of total stockholdings are needed (see, e.g., Fama and French, 2007 for statistics on migration rates for high- and low-B/P stocks). However, there are differences in the needs for rebalancing trades between the portfolios, as their turnover rates vary across the portfolio-formation criteria, as well as over time. Generally, the inclusion of transaction costs would slightly decrease the statistical significances of the outperformance of the best portfolios while increasing the significances of the underperformance of the worst. In addition, the inclusion of transaction costs would most probably favor the portfolios that consist of larger-cap stocks, since the price impact from implementing trades is stronger among smaller-cap stocks (see, e.g., Chiyachantana *et al.*, 2004).

7 Conclusions

This paper compares the efficacy of both traditional valuation ratios and a comprehensive set of related combination criteria in identifying the best- and worst-performing stocks of the future for a comprehensive sample of U.S. stocks over a 42-year period from May 1971 to April 2013. The results for the single selection criteria show that the pricing anomalies based on the most commonly used valuation ratios, such as E/P and B/P, are somewhat weaker than pricing anomalies based on many less frequently used, or less examined, valuation ratios. We find evidence of strong efficacy for the three EV-based valuation ratios (EBIT/EV, EBITDA/EV, and S/EV). In particular, the evidence for the unique characteristics of S/EV contributes to the existing literature. Altogether, our overall results are in line with McLean and Pontiff (2016) who conclude that pricing anomalies tend to attenuate after their publication. On the other hand, even the first-detected value anomalies have not totally disappeared during the 42-year sample period, which supports both the Adaptive Markets Hypothesis of Lo (2004) and the theory of gradual decay of anomalies introduced by Jones and Pomorski (2017).

We also show that the dividend yield criterion has been particularly useful for risk reduction purposes to the extent that, in terms of the Sharpe ratio, the highest D/P decile portfolio is the best among all the decile portfolios formed on the 12 individual valuation ratios examined. The low-risk characteristic of the top-decile D/P portfolio also extends to the combination criteria in which D/P is included as one sub-criterion. Moreover, all the high-D/P decile portfolios are tilted towards larger-cap stocks, which also extends to the D/P-related combinations. Overall, we find remarkable differences between the valuation ratios with respect to the market-cap tilts of the related value portfolios.

Appendix

Table 1A Average regression coefficients of decile portfolios for the 18 selection criteria.

Decile portfolio	Alpha (%)		MKT		SMB		HML		WML		SMV	
Panel A: *Decile averages for 5-factor regression coefficients*												
D1	1.67	(1.23)	0.977	(1.002)	−0.017	(0.043)	0.426	(0.436)	−0.097	(−0.113)	0.043	(0.051)
D2	0.33	(0.32)	0.992	(0.991)	0.017	(0.061)	0.324	(0.345)	−0.061	(−0.071)	0.095	(0.110)
D3	0.03	(0.07)	0.991	(0.996)	0.029	(0.032)	0.254	(0.275)	−0.055	(−0.061)	0.113	(0.118)
D4	−0.54	(−0.31)	1.003	(1.003)	0.009	(−0.007)	0.187	(0.215)	−0.038	(−0.036)	0.098	(0.103)
D5	−0.34	(−0.45)	0.994	(0.997)	−0.047	(−0.048)	0.069	(0.089)	−0.009	(−0.004)	0.038	(0.064)
D6	−0.69	(−1.11)	0.996	(0.986)	−0.040	(−0.053)	−0.031	(−0.032)	0.014	(0.022)	0.014	(0.036)
D7	−0.55	(−0.37)	1.010	(1.007)	−0.032	(−0.051)	−0.124	(−0.150)	0.024	(0.035)	−0.002	(−0.020)
D8	−0.28	(−0.09)	1.012	(1.015)	−0.009	(−0.017)	−0.248	(−0.264)	0.037	(0.035)	−0.041	(−0.045)
D9	−0.31	(−0.28)	1.015	(1.001)	−0.037	(−0.090)	−0.372	(−0.415)	0.056	(0.059)	−0.105	(−0.106)
D10	−0.04	(0.15)	1.020	(1.022)	0.047	(0.135)	−0.383	(−0.389)	0.110	(0.110)	−0.172	(−0.184)
Panel B: *Summary statistics*												
No. of sign. cases at (5% level)	26		180		79		154		101		140	
+/−	10/16		180/0		36/43		83/71		42/59		91/49	
min	−2.66		0.855		−0.413		−0.719		−0.171		−0.341	
max	4.11		1.109		0.259		0.545		0.236		0.188	

Panel A presents average 5-factor alphas (p.a.) and the corresponding regression slopes for the decile portfolios formed on the basis of 18 different selection criteria, of which 12 are individual valuation ratios and the remaining six are the top combination criteria that generate either the highest raw returns or the highest Sharpe ratio (as well as the highest SKASR) within 2-, 3-, or 4-combinations. The corresponding medians are given in parentheses. Alphas are calculated according to Eq. (1). Panel B indicates the total number of significant regression coefficients, as well as the numbers of significantly positive/negative coefficients (+/−). The two bottom rows indicate the minimum and maximum values within all 180 decile portfolios included in this analysis.

Table 2A 5-factor regression results for the top- and bottom-decile portfolios.

Portfolio-formation criterion	Alpha (%)	p (%)	MKT	p (%)	SMB	p (%)	HML	p (%)	WML	p (%)	SMV	p (%)	R^2
Panel A: 5-factor regression results for top-decile portfolios													
S/EV	4.11	(0.0)	0.924	(0.0)	0.154	(0.8)	0.173	(0.0)	0.030	(42.0)	−0.159	(0.0)	0.924
EBIT/EV[a]	3.19	(0.0)	0.968	(0.0)	0.148	(0.0)	0.306	(0.0)	−0.115	(0.0)	0.049	(2.2)	0.943
EBITDA/EV	2.95	(0.1)	0.961	(0.0)	0.090	(2.7)	0.376	(0.0)	−0.083	(0.0)	0.006	(78.4)	0.942
MS2 (CF3/P, S/EV)[a]	2.93	(0.1)	1.004	(0.0)	0.044	(38.9)	0.450	(0.0)	−0.106	(0.0)	0.021	(43.4)	0.940
MS4 (B/P, CF1/P, EBITDA/EV, S/EV)[a]	2.92	(0.1)	0.999	(0.0)	0.042	(42.4)	0.509	(0.0)	−0.145	(0.0)	0.026	(28.8)	0.941
MS3 (EBITDA/EV, S/EV, S/P)[a]	2.79	(0.5)	1.031	(0.0)	0.060	(30.3)	0.484	(0.0)	−0.138	(0.0)	0.040	(14.7)	0.931
S/P	1.88	(4.8)	1.035	(0.0)	0.086	(17.9)	0.515	(0.0)	−0.151	(0.0)	0.046	(9.2)	0.930
MS4 (CFO/P, D/P, EBITDA/EV, S/EV)[b]	1.28	(18.0)	0.904	(0.0)	−0.350	(0.0)	0.335	(0.0)	−0.037	(24.7)	0.084	(0.1)	0.879
MS3 (D/P, EBITDA/EV, S/EV)[b]	1.27	(18.4)	0.894	(0.0)	−0.347	(0.0)	0.341	(0.0)	−0.030	(33.5)	0.076	(0.3)	0.876
CFO/P	1.19	(14.9)	1.010	(0.0)	0.026	(62.0)	0.480	(0.0)	−0.123	(0.0)	0.049	(5.3)	0.933
B/P	1.08	(18.3)	1.011	(0.0)	0.127	(1.5)	0.542	(0.0)	−0.171	(0.0)	0.017	(46.6)	0.947
MS2 (D/P, EBIT/EV)[b]	0.94	(31.7)	0.867	(0.0)	−0.357	(0.0)	0.360	(0.0)	−0.040	(19.0)	0.083	(0.1)	0.870
CF/P	0.94	(26.2)	1.036	(0.0)	−0.030	(60.3)	0.544	(0.0)	−0.116	(0.0)	0.052	(4.6)	0.935
CF3/P	0.92	(26.0)	1.043	(0.0)	−0.036	(52.2)	0.544	(0.0)	−0.111	(0.0)	0.057	(2.4)	0.938
E/P	0.59	(51.1)	1.006	(0.0)	0.052	(32.4)	0.418	(0.0)	−0.159	(0.0)	0.066	(1.3)	0.926
CF2/P	0.58	(48.7)	1.046	(0.0)	−0.014	(80.0)	0.545	(0.0)	−0.122	(0.0)	0.061	(1.6)	0.935
FCF/EV	0.30	(66.7)	0.991	(0.0)	0.070	(2.4)	0.333	(0.0)	−0.063	(0.0)	0.056	(0.2)	0.958
D/P[b]	0.21	(78.6)	0.855	(0.0)	−0.074	(16.3)	0.421	(0.0)	−0.069	(0.1)	0.147	(0.0)	0.899

This table presents 5-factor alphas (p.a.) and the corresponding regression slopes for the top- and bottom-decile portfolios (in Panels A and B, respectively) formed on the 12 individual valuation ratios and the top combination criteria that generate either the highest raw returns or the highest Sharpe ratio within 2-, 3-, or 4-combinations (significances in percentages are in parentheses). Alphas are calculated according to Eq. (1). The first column indicates the selection criterion, whereas the next columns indicate the alphas, the regression slopes for market return, SMB, HML, and momentum factors, respectively. The last column indicates the R-squareds of the regressions for each of the decile portfolios. The selection criteria are in descending order according to the 5-factor alphas of the top-decile portfolios.

[a]Indicates that the criterion has generated the highest top-decile return among either 2-, 3-, or 4-combinations, or among individual valuation ratios.

[b]Indicates the corresponding superiority in terms of the Sharpe ratio and the SKASR.

Table 2A (*Continued*)

Portfolio-formation criterion	alpha (%)	*p* (%)	MKT	*p* (%)	SMB	*p* (%)	HML	*p* (%)	WML	*p* (%)	SMV	*p* (%)	R^2
Panel B: *5-factor regression results for bottom-decile portfolios*													
S/EV	−2.48	(5.6)	1.045	(0.0)	−0.413	(0.0)	−0.091	(6.9)	0.107	(0.5)	0.002	(96.4)	0.861
EBIT/EV[a]	1.44	(37.4)	0.946	(0.0)	0.146	(24.2)	−0.370	(0.0)	0.141	(2.3)	−0.341	(0.0)	0.895
EBITDA/EV	0.95	(55.4)	0.956	(0.0)	0.151	(22.7)	−0.596	(0.0)	0.149	(1.5)	−0.299	(0.0)	0.896
MS2 (CF3/P, S/EV)[a]	0.17	(90.8)	1.027	(0.0)	−0.012	(91.8)	−0.407	(0.0)	0.096	(7.0)	−0.179	(0.0)	0.894
MS4 (B/P, CF1/P, EBITDA/EV, S/EV)[a]	0.14	(92.6)	1.016	(0.0)	0.020	(85.9)	−0.587	(0.0)	0.120	(2.1)	−0.190	(0.0)	0.898
MS3 (EBITDA/EV, S/EV, S/P)[a]	−0.42	(77.1)	0.972	(0.0)	−0.028	(82.4)	−0.590	(0.0)	0.112	(3.9)	−0.176	(0.0)	0.887
S/P	0.75	(62.2)	0.908	(0.0)	−0.067	(54.4)	−0.719	(0.0)	0.236	(0.0)	−0.284	(0.0)	0.898
MS4 (CFO/P, D/P, EBITDA/EV, S/EV)[b]	−0.97	(48.2)	1.081	(0.0)	0.149	(1.4)	−0.295	(0.0)	0.086	(1.6)	−0.077	(2.0)	0.914
MS3 (D/P, EBITDA/EV, S/EV)[b]	−0.64	(64.3)	1.086	(0.0)	0.140	(3.0)	−0.340	(0.0)	0.082	(2.1)	−0.075	(3.1)	0.914
CFO/P	0.42	(79.1)	1.033	(0.0)	0.160	(21.2)	−0.321	(0.0)	0.108	(8.4)	−0.208	(0.0)	0.902
B/P	1.17	(28.8)	0.986	(0.0)	−0.129	(3.9)	−0.635	(0.0)	0.212	(0.0)	−0.163	(0.0)	0.939
MS2 (D/P, EBIT/EV)[b]	−0.13	(92.4)	1.077	(0.0)	0.237	(0.0)	−0.227	(0.0)	0.089	(1.1)	−0.104	(0.1)	0.919
CF/P	0.54	(73.2)	1.017	(0.0)	0.140	(26.2)	−0.479	(0.0)	0.115	(5.3)	−0.221	(0.0)	0.898
CF3/P	0.54	(72.8)	0.999	(0.0)	0.155	(19.6)	−0.450	(0.0)	0.125	(3.4)	−0.231	(0.0)	0.903
E/P	0.12	(93.7)	1.060	(0.0)	0.082	(37.1)	−0.137	(2.0)	0.076	(14.2)	−0.205	(0.0)	0.913
CF2/P	0.95	(54.8)	1.003	(0.0)	0.130	(30.5)	−0.461	(0.0)	0.125	(3.7)	−0.242	(0.0)	0.900
FCF/EV	−1.29	(32.2)	1.050	(0.0)	0.174	(0.3)	−0.076	(8.8)	−0.051	(30.6)	−0.123	(0.2)	0.930
D/P[b]	−2.06	(6.0)	1.095	(0.0)	−0.186	(0.3)	−0.117	(0.9)	0.057	(9.0)	0.020	(50.0)	0.909

Acknowledgements

We thank an anonymous referee, as well as Jukka Perttunen and other seminar participants at the Graduate School of Finance Summer Workshop 2016 at the University of Vaasa for useful comments and suggestions. We are also grateful to Teemu Toivonen for excellent research assistance.

Notes

1 E.g., Jaffe *et al.* (1989) found evidence of high returns for U.S. firms of all sizes with negative earnings. For an extreme example, Chan *et al.* (1991) documented the highest returns for the portfolio that included only firms with negative earnings in the Japanese stock markets. The parallel international evidence was given by Bauman *et al.* (1998).

2 Although S/EV has been included as one sub-criterion in the composite value criteria in few recent studies (see, e.g., Israel and Maloney, 2014; Ilmanen *et al.*, 2015), we are not aware of any scientific article in which S/EV would have been compared as a stand-alone valuation criterion with other individual value multiples or in which it would have been systematically combined with varying sets of other sub-criteria.

3 E.g., Pätäri *et al.* (2016) show that leverage-adjustment enhanced the efficacy of B/P and E/P ratios for value portfolio selection (see also Palkar and Wilcox, 2009; Leibowitz, 2002, for a comprehensive analysis of the relationship between E/P ratios and leverage).

4 In exceptional cases, it may happen that, as a result of higher leverage, the stock price may rise by a higher percentage than that of sales growth generated by increased debt. However, in most cases, the impact of increased debt on the S/P ratio is positive (see, e.g., Damodaran, 2012 for details).

5 We use estimates only in 0.65% of cases, which implies that our sample is essentially free of delisting bias.

6 For Equity's Share of Depreciation, we adopt the following definition of Fama and French (2006a): ME/(Total Assets (item AT)−BE (item SEQ) + ME)*Depreciation and Amortization (item DP).

7 Accruals are computed according to the definition of Sloan (1996) as follows: [ΔCurrent Assets (item ACT) − ΔCash (item CH)] − [ΔCurrent Liabilities (item LCT) − ΔShort-term Debt (item DLC) − ΔTax Payable (item TXP)] − Depreciation and Amortization (item DP), where Δ represents the annual change.

8 As a robustness check, we recalculated the decile returns for all individual valuation ratios without such an exclusive criterion, and found that it had only marginal impact on the results (see Subsection 5.3. for details).

9 Because the availability of accounting data decreases dramatically when moving back in time to the 1960s, we decided to limit our sample period to 1971–2013 to ensure that even the narrowest decile portfolios at the beginning of the sample period are properly diversified.

10 Like Jegadeesh and Titman (2001), Avramov *et al.* (2007), and Fong and Ong (2016), we also set the penny stock filter at $5, but found its impact negligible, thus indicating that almost all stocks that have traded below $5 at each checkpoint had already been filtered out of the sample on the basis of some other exclusion criteria.

11 We also tested 5- and 6-combinations but their marginal utility compared to narrower combinations was mostly negative; therefore, we restricted our analysis to 2-, 3-, and 4-combinations.

12 In the case of D/P, the cross-sectional medians are determined from the sub-sample of dividend-paying stocks because for many years, the median D/P for the full sample is zero.

13 This methodology is followed instead of using value-weighted returns because from the viewpoint of practical portfolio management, which is the focus of this study, our approach is more realistic (given that value-weighting attenuates value anomalies (see, e.g., Arnott and Hsu, 2008; Loughran and Wellman, 2011), hardly any portfolio manager aiming to exploit these anomalies would weight the constituent stocks based on their market caps when deciding on equity portfolio allocation. See also the discussion in Fama and French (2008) on pros and cons of the use of value-weighted returns in anomaly studies).

14 In addition to the standard Sharpe ratios, we calculate the skewness- and kurtosis-adjusted Sharpe ratios (SKASR) to control for the impact of the violation of the hypothesis on normally-distributed returns. The skewness- and kurtosis-adjustment is made by multiplying the standard deviation in the denominator of the standard Sharpe ratio by the ratio Z_{CF}/Z_c, where Z_c is the critical value of the probability based on the standard normal distribution (set to -1.96 to correspond to the 95% probability level), and Z_{CF} is the corresponding skewness- and kurtosis-adjusted value calculated on the basis of the fourth-order Cornish–Fisher expansion as follows:

$$Z_{CF} = Z_C + (Z_C^2 - 1)S/6 + (Z_C^3 - 3Z_C)K/24 - (2Z_C^3 - 5Z_C)S^2/36,$$

where S refers to Fisher's skewness and K to excess kurtosis (see Pätäri, 2011, for the introduction of the SKASR and the related risk metrics *skewness- and kurtosis-adjusted deviation* (SKAD)).

15 For this particular sample, all those 2-, 3-, and 4-combination portfolios that perform best in terms of the Sharpe ratio are also the best based on the SKASR.

16 We use the sample average return instead of the value-weighted stock market return employed in the standard Fama–French factor models because the former is a more valid proxy for the market portfolio return in cases where the performance of decile portfolios is benchmarked against the sample average return. This also makes the 5-factor alphas more comparable to the total risk-based performance metrics, because the significance levels for outperformance/underperformance based on the latter are calculated against the equal-weighted time-series of sample average returns. (This choice is also reasonable because the equal-weighted average return is clearly higher than the corresponding value-weighted market return employed in the standard Fama–French factor models. For example, for our sample, the equal-weighted geometric average annual return from May 1971 to April 2013 is approximately 13.0% p.a., whereas the corresponding value-weighted Fama–French stock market return is 10.1%. The return difference is partially explained by the well-documented size anomaly (for recent evidence of this, see e.g., Israel and Moskowitz, 2013) and partially by lower average returns of stocks with negative book equity (according to French's data library, the corresponding equal-weighted return for such stocks has been only 6.9% p.a. during the sample period)).

17 Fama and French (2016) showed that nothing is lost in terms of model performance if HML is dropped from their 5-factor model, because other factors capture the information in HML about average returns.

18 The RMW factor is calculated based on the return difference between robust and weak operating profitability stocks (see Fama and French, 2016 for details).

19 We suppose that this may stem from the failure of such a 5-factor model to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability, as put forth by Fama and French (2015). The results were also similar for the 6-factor model employed indicating that the observed U-shaped relation is not explained by the omission of momentum in the Fama–French 5-factor model. However, when RMW was dropped from the 6-factor model, the results became very similar to those indicated by the Carhart 4-factor model.

20 We are grateful to an anonymous referee for suggesting the inclusion of the volatility factor. We tested several variants of volatility factors, including the corresponding beta-based and idiosyncratic volatility-based factors, but ended up using the one based on the basic volatility, as it produced the greatest improvement in the average-adjusted R-squared, as well as in the proportion of significant regression coefficients for that specific factor. However, the results were qualitatively the same for all the volatility factors being tested.

21 Because of the complexity of the test procedure and space limitations we do not describe the Ledoit–Wolf test in detail here, but recommend the interested reader see the original article (Ledoit and Wolf, 2008. The corresponding programming code is freely available at: http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9).

22 Among the five explanatory factors, the lowest tolerance (i.e., the highest multicollinearity) is documented for the volatility factor, for which it is 29.3% (tolerance measures how much variance in an explanatory factor is unique to that factor or not related to or explained by the other explanatory factors).

23 Because the average proportion of non-dividend-paying stocks for this sample is 41.8% over the 42-year sample period, we calculate the top-2 decile returns formed on D/P based on both the sample that includes only dividend-paying stocks and the sample that also includes zero-dividend stocks. The results for the top-2 deciles reported in the tables are for the latter sample, since, for the former sample, the D/P decile portfolios are much narrower than their counterpart portfolios formed on other selection criteria, which would have diminished the comparability of the D/P-based results with those based on other selection criteria (However, the differences between the results of these two samples are quite marginal, although the former (narrower) sample generates a somewhat higher average return (15.43% vs. 15.04%) for the D/P-based top deciles. Nonetheless, the Sharpe ratios of these two top-decile portfolios are almost equal within the samples (0.623 and 0.621, respectively). The returns of the other D/P decile portfolios were calculated only for the sample of dividend-paying stocks because of the high proportion of zero-dividend stocks (which is above the bottom-decile breakpoint in all years included in the sample

period, and even above the bottom quintile breakpoint in 36 out of 42 years). Therefore, it would have been impossible to divide zero-dividend stocks into the lowest deciles if the aim had been to get equal-sized portfolios. For the same reason, the decile return average for the reported D/P deciles deviates from that of other selection criteria (the average return is higher for dividend-paying stocks than for zero-dividend stocks, for which it is 12.65% p.a.; the corresponding Sharpe ratio is 0.272). However, all the other decile portfolios (including those formed on the basis of combinations in which D/P is one of the selection criteria) are of equal size, since in such cases, the other selection criterion/criteria enable(s) the ranking of all stocks, including zero-dividend ones.

24 Because some valuation ratios are close proxies for each other, we restricted the number of potential combinations so that only one of the price-based earnings or cash flow multiples (i.e., E/P, CF1/P, CF2/P, CF3/P, or CFO/P) and one of the corresponding EV-based multiples (i.e., EBIT/EV, EBITDA/EV, or FCF/EV) can occur at time in the combinations when calculating the averages for 2-, 3-, and 4-combinations reported in Table 3. This implies that the reported averages are based on fifty-three 2-combinations, one hundred twelve 3-combinations, and one hundred twenty-three 4-combinations (We also calculated the averages without any restrictive conditions. The results (available upon request) are very similar to those reported in Table 3).

25 Interestingly, for such 2-combination (3-combination), the fifth (sixth) decile alpha is positive and significant.

26 All the rank correlation relations remained qualitatively the same when the Sharpe ratio rankings were replaced with the SKASR rankings.

27 On average, the negative impact of higher negative skewness of the latter-sub-period market returns on the skewness- and kurtosis-adjusted performance statistics is approximately fully compensated by lower kurtosis (compared to that of the market return distribution for the first sub-period).

28 As a supplementary robustness check for the dependency of relative performance of different decile portfolios on the stock market trend, we also divided the sample period into bull and bear market periods according to the turning points of the U.S. stock market, following Edwards *et al.* (2003), who used a 20% cumulative return (loss) from the previous trough (peak) to the subsequent peak (trough) in the demarcation of bullish (bearish) periods. The un-tabulated results are qualitatively the same and reinforce the superiority of the top-decile D/P portfolio during the bearish periods.

29 For the sample data employed, we also calculated the monthly factor returns based on each of the 12 valuation ratios as the return differences between the top-30% and the bottom-30% of the sample stocks. The correlation analysis between these factors and Fama–French HML and WML factors revealed that the S/EV factor behaved differently to any other value factors, as it did not significantly correlate with either of the HML and WML factors. By contrast, the value factors formed on the other 11 valuation ratios all correlated positively and significantly with HML, and 10 of them also correlated negatively and significantly with WML (the exception was the FCF/EV that had zero-correlation with WML).

30 Although this conclusion cannot be directly drawn from the separate analyses of the bullish and bearish months, the results of un-tabulated additional analysis of bullish and bearish periods described in Endnote 28 prove that this generalization also holds for this particular sample data.

31 With respect to low-volatility characteristic of the top-decile D/P portfolios, our results are generally consistent with Naranjo *et al.* (1998) who also report the lowest volatility for the same D/P decile. However, it is noteworthy that due to differences in samples and methodologies, their results are not directly comparable to ours. For example, Naranjo *et al.* (1998) restrict their analysis to NYSE firms with an established track record of quarterly dividend payments, and calculate the dividend yield by multiplying a firm's most recently declared quarterly dividend by four and dividing the resulting product by the previous month's closing price (see the original article for details). Recently, Fong and Ong (2016) report the similar characteristic for the value-weighted top-quintile D/P portfolio.

32 In this respect, our results are in line with those of Fong and Ong (2016) for the value-weighted quintile portfolios.

33 On the other hand, the investor could neither have identified the best-performing individual valuation ratios beforehand.

34 With respect to more sophisticated combination methods, Eakins and Stansell (2003) use neural networking procedures for the selection of value portfolios, whereas Pätäri *et al.* (2010, 2012) employ data envelopment analysis for the same purpose. As an example of a

simpler combination method, Israel and Moskowitz (2013) allocate the stocks into portfolios in accordance with their average rankings based on several individual valuation ratios.

References

Arnott, R. and Hsu, J. (2008). "Noise, CAPM and the Size and Value Effects," *Journal of Investment Management* **6**, 1–11.

Asness, C. S., Ilmanen, A., Israel, R., and Moskowitz, T. J. (2015). "Investing with Style," *Journal of Investment Management* **13**, 27–63.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). "Value and Momentum Everywhere," *Journal of Finance* **68**, 929–985.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2007). "Momentum and Credit Rating," *Journal of Finance* **62**, 2503–2520.

Barbee, W. Jr., Jeong, J., and Mukherji, S. (2008). "Relations Between Portfolio Returns and Market Multiples," *Global Finance Journal* **19**, 1–10.

Bartov, E. and Kim, M. (2004). "Risk, Mispricing, and Value Investing," *Review of Quantitative Finance and Accounting* **23**, 353–376.

Bauman, S. W., Conover, M. C., and Miller, R. E. (1998). "Growth versus Value and Large-Cap versus Small-Cap Stocks in International Markets," *Financial Analysts Journal* **54**, 75–89.

Beaver, W., McNichols, M., and Price, R. (2007). "Delisting Returns and Their Effect on Accounting-Based Market Anomalies," *Journal of Accounting and Economics* **43**, 341–368.

Bird, R. and Casavecchia, L. (2007). "Value Enhancement Using Momentum Indicators: The European Experience," *International Journal of Managerial Finance* **3**, 229–262.

Bird, R. and Whitaker, J. (2003). "The Performance of Value and Momentum Investment Portfolios: Recent Experience in the Major European Markets," *Journal of Asset Management* **4**, 221–246.

Bodie, Z., Kane, A., and Marcus, A. J. (2010). *Investments*. (9th ed.). Singapore: McGraw-Hill/Irwin, (Chapter 24).

Cakici, N., Fabozzi, F. J., and Tan, S. (2013). "Size, Value, and Momentum in Emerging Market Stock Returns," *Emerging Markets Review* **16**, 46–65.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance," *Journal of Finance* **52**, 57–82.

Chan, L. K. C., Hamao, Y., and Lakonishok, J. (1991). "Fundamentals and Stock Returns in Japan," *Journal of Finance* **46**, 1739–1765.

Chan, L. and Lakonishok, J. (2004). "Value and Growth Investing: Review and Update," *Financial Analysts Journal* **60**, 71–86.

Chiyachantana, C. N., Jain, P. K., Jiang, C., and Wood, R. A. (2004). "International Evidence on Institutional Trading Behavior and Price Impact," *Journal of Finance* **59**, 869–898.

Clarke, R., de Silva, H., and Thorley, S. (2010). "Know Your VMS Exposure," *Journal of Portfolio Management* **36**, 52–59.

Clarke, R., de Silva, H., and Thorley, S. (2014). "The Not-So-Well-Known Three-and-One-Half-Factor Model," *Financial Analysts Journal* **70**, 13–23.

Damodaran, A. (2012). *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*. (3rd ed.). Hoboken, NJ: John Wiley & Sons (Chapter 20).

Davis, J. L. (1994). "The Cross-Section of Realized Stock Returns: The Pre-Compustat Evidence," *Journal of Finance* **49**, 1579–1593.

Davis, J. L., Fama, E. F., and French, K. R. (2000). "Characteristics, Covariances, and Average Returns: 1929 to 1997," *Journal of Finance* **55**, 389–406.

Desai, H., Rajgopal, S., and Venkatachalam, M. (2004). "Value-Glamour and Accruals Mispricing: One Anomaly or Two?" *Accounting Review* **79**, 355–385.

Dhatt, M. S., Kim, Y. J., and Mukherji, S. (2004). "Can Composite Value Measures Enhance Portfolio Performance?" *Journal of Investing* **13**, 42–48.

Eakins, S. G. and Stansell, S. R. (2003). "Can Value-Based Stock Selection Criteria Yield Superior Risk-Adjusted Returns: An Application of Neural Networks," *International Review of Financial Analysis* **12**, 83–97.

Edwards, S., Biscarri, J. G. and de Gracia, F. P. (2003) "Stock Market Cycles, Financial Liberalization and Volatility," *Journal of International Money and Finance* **22**, 925–955.

Fama, E. F. and French, K. R. (1992). "The Cross-Section of Expected Stock Returns," *Journal of Finance* **47**, 427–465.

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

Fama, E. F. and French, K. R. (1998). "Value versus Growth: The International Evidence," *Journal of Finance* **53**, 1975–1999.

Fama, E. F. and French, K. R. (2001). "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics* **60**, 3–43.

Fama, E. F. and French, K. R. (2006a). "Profitability, Investment and Average Return," *Journal of Financial Economics* **82**, 491–518.

Fama, E. F. and French, K. R. (2006b). "The Value Premium and the CAPM," *Journal of Finance* **61**, 2163–2185.

Fama, E. F. and French, K. R. (2007). "Migration," *Financial Analysts Journal* **63**, 48–58.

Fama, E. F. and French, K. R. (2008). "Dissecting Anomalies," *Journal of Finance* **63**, 1653–1678.

Fama, E. F. and French, K. R. (2012). "Size, Value, and Momentum in International Stock Returns," *Journal of Financial Economics* **105**, 457–472.

Fama, E. F. and French, K. R. (2015). "A Five-Factor Asset Pricing Model," *Journal of Financial Economics* **116**, 1–22.

Fama, E. F. and French, K. R. (2016). "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies* **29**, 69–103.

Fisher, G., Shah, R., and Titman, S. (2016). "Combining Value and Momentum," *Journal of Investment Management* **14**, 33–48.

Fong, W. M. and Ong, Z. (2016). "The Profitable Dividend Yield Strategy," *Journal of Investment Management* **14**, 51–61.

Frazzini, A., Israel, R., and Moskowitz, T. (2015). "Trading Costs of Asset Pricing Anomalies," Unpublished working paper. University of Chicago, IL.

French, K. R., Data library. (2015). http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html/ Accessed 15.12.20.

Fuller, K. P. and Goldstein, M. A. (2011). "Do Dividends Matter More in Declining Markets?" *Journal of Corporate Finance* **17**, 457–473.

Gray, W. R. and Vogel, J. (2012). "Analyzing Valuation Measures: A Performance Horse-Race Over the Past 40 years," *Journal of Portfolio Management* **39**, 112–121.

Hou, K., Karolyi, G. A., and Kho, B. C. (2011). "What Factors Drive Global Stock Returns?," *Review of Financial Studies* **24**, 2527–2574.

Hou, K., Xue, C., and Zhang, L. (2015). "Digesting Anomalies: An Investment Approach," *Review of Financial Studies* **28**, 650–705.

Hwang, S. and Rubesam, A. (2013). "A Behavioral Explanation of the Value Anomaly based on Time-Varying Return Reversals," *Journal of Banking and Finance* **37**, 2367–2377.

Ilmanen, A., Nielsen, L. N., and Chandra, S. (2015). "Are Defensive Stocks Expensive? A Closer Look at Value Spreads," AQR white paper.

Israel, R. and Maloney, T. (2014). "Understanding Style Premia," *Journal of Investing* **23**, 15–22.

Israel, R. and Moskowitz, T. J. (2013). "The Role of Shorting, Firm Size, and Time on Market Anomalies," *Journal of Financial Economics* **108**, 275–301.

Jaffe, J., Keim, D. B., and Westerfield, R. (1989). "Earnings Yield, Market Values and Stock Returns," *Journal of Finance* **44**, 135–148.

Jarque, C. M. and Bera, A. K. (1980). "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economics Letters* **6**, 255–259.

Jegadeesh, N. and Titman, S. (2001). "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *Journal of Finance* **56**, 699–720.

Jones, C. S. and Pomorski, L. (2017). "Investing in Disappearing Anomalies," *Review of Finance* **21**, 237–267.

Keim, D. B. and Madhavan, A. (1997). "Transactions Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades," *Journal of Financial Economics* **46**, 265–292.

Kosowski, R., Naik, N. Y., and Teo, M. (2007). "Do Hedge Funds Deliver Alpha? A Bayesian and Bootstrap Analysis," *Journal of Financial Economics* **84**, 229–264.

Lakonishok, J., Shleifer, A., and Vishny, R. (1994). "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance* **49**, 1541–1578.

Ledoit, O. and Wolf, M. (2008). "Robust Performance Hypothesis Testing with the Sharpe Ratio," *Journal of Empirical Finance* **15**, 850–859.

Leibowitz, M. (2002). "The Levered P/E Ratio," *Financial Analysts Journal* **58**, 68–77.

Lewellen, J. (2010). "Accounting Anomalies and Fundamental Analysis: An Alternative View," *Journal of Accounting and Economics* **50**, 455–466.

Li, X., Brooks, C., and Miffre, J. (2009). "The Value Premium and Time-Varying Volatility," *Journal of Business Finance and Accounting* **36**, 1252–1272.

Lo, A. W. (2004). "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective," *Journal of Portfolio Management* **30**, 15–29.

Loughran, T. and Wellman, J. (2011). "New Evidence on the Relation between the Enterprise Multiple and Average Stock Returns," *Journal of Financial and Quantitative Analysis* **46**, 1629–1650.

McLean, R. D. and Pontiff, J. (2016). "Does Academic Research Destroy Stock Return Predictability?," *Journal of Finance* **71**, 5–32.

Newey, W. K. and West, K. D. (1987). "A Simple Semidefinite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* **55**, 703–708.

Naranjo, A., Nimalendran, M., and Ryngaert, M. (1998). "Stock Returns, Dividend Yields, and Taxes," *Journal of Finance* **53**, 2029–2057.

Novy-Marx, R. (2013). "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics* **108**, 1–28.

Novy-Marx, R. and Velikov, M. (2016). "A Taxonomy of Anomalies and Their Trading Costs," *Review of Financial Studies* **29**, 104–147.

Palkar, D. D. and Wilcox, S. E. (2009). "Adjusted Earnings Yields and Real Rates of Return," *Financial Analysts Journal* **65**, 66–79.

Penman, S. H. (2013). *Financial Statement Analysis and Security Valuation.* (5th ed). Singapore: McGraw-Hill, (Chapter 3).

Pätäri, E. J. (2011). "Does the Risk-Adjustment Method Matter at All in Hedge Fund Rankings?," *International Research Journal of Finance and Economics* **6**, 69–99.

Pätäri, E. J. and Leivo, T. H. (2017). "A Closer Look at Value Premium: Literature Review and Synthesis," *Journal of Economic Surveys* **31**, 79–168.

Pätäri, E. J., Leivo, T. H., and Honkapuro, J. V. S. (2010). "Enhancement of Value Portfolio Performance Using Data Envelopment Analysis," *Studies in Economics and Finance* **27**, 223–246.

Pätäri, E. J., Leivo, T. H., and Honkapuro, J. V. S. (2012). "Enhancement of Equity Portfolio Performance Using Data Envelopment Analysis," *European Journal of Operational Research* **220**, 786–797.

Pätäri, E. J., Karell, V., and Luukka, P. (2016). "Can Size-, Industry-, and Leverage-Adjustment of Valuation Ratios Benefit the Value Investor?," *International Journal of Business Innovation and Research* **11**, 76–109.

Sharpe, W. F. (1966). "Mutual Fund Performance," *Journal of Business* **39**, 119–138.

Simlai, P. E. (2016). "Time-Varying Risk, Mispricing Attributes, and the Accrual Premium," *International Review of Financial Analysis* **48**, 150–161.

Sloan, R. G. (1996). "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *Accounting Review* **71**, 289–315.

Strydom, M., Skully, M., and Veeraraghavan, M. (2014). "Is the Accrual Anomaly Robust to Firm-Level Analysis?," *International Review of Financial Analysis* **34**, 157–165.

Treynor, J. L. and Black, F. (1973). "How to Use Security Analysis to Improve Portfolio Selection," *Journal of Business* **46**, 66–86.

Walkshäusl, C. and Lobe, S. (2015). "The Enterprise Multiple Investment Strategy: International Evidence," *Journal of Financial and Quantitative Analysis* **50**, 781–800.

Keywords: Value anomalies; value investing; portfolio performance; value premium; valuation ratios