European Journal of Operational Research 265 (2018) 655–672

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Comparison of the multicriteria decision-making methods for equity portfolio selection: The U.S. evidence

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

[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

ARTICLE INFO

Article history:
Received 23 March 2017
Accepted 1 August 2017
Available online 12 August 2017

Keywords:
Anomalies
Value investing
Momentum
Multicriteria decision-making

ABSTRACT

This paper compares the efficacy of four multicriteria decision-making (MCDM) methods in identifying the future best-performing stocks in two comprehensive samples of U.S. stocks. This is the first time that median-scaling (MS), the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), the Analytic Hierarchy Process (AHP), and the additive Data Envelopment Analysis (add.DEA) have been used to combine value and momentum indicators into a single efficiency score. The results show that the MCDM methods examined can successfully be applied to equity portfolio selection. As a robustness check, we repeat all the main sample tests for the sample of the largest-cap stocks included in the two biggest size quintiles (i.e., stocks above 40% NYSE market-cap breakpoint) and find that the overall results are surprisingly robust to size effect. However, the best-performing portfolios formed on the basis of different MCDM methods have remarkably different exposures to the style factors that are commonly used to explain the abnormal returns of active equity portfolios. As a practical implication of this study, investors following certain investing styles could take these different style exposures into account when choosing the MCDM criteria that best fit their portfolio-selection purposes.

1. Introduction

Considerable evidence against the efficient market hypothesis has been documented over the past four decades. The extensive literature on value anomalies has shown that value stocks tend to outperform glamour stocks, as well as the market portfolio, most of the time[1] (see, e.g., Chan & Lakonishok, 2004; Cakici, Fabozzi, & Tan, 2013; Fama & French, 2006b, 2012). Evidence has shown not only that the value anomalies in stock markets are a worldwide phenomenon, but also that the relative efficacy of different valuation criteria varies across both the stock markets and the sample periods examined (for a comprehensive literature review on value anomalies, see Pätäri & Leivo, 2017). Because of the diversity of the results, it might be worthwhile to combine several individual valuation ratios into a single composite value criterion to create more stable portfolios.

Another branch of financial literature has identified the existence of price momentum in stock returns, which refers to the tendency of recent winner stocks to generate abnormal returns also in the near future (see, e.g., Barroso & Santa-Clara, 2015; Billio, Calès, & Guégan, 2011; Chan, Hameed, & Tong, 2000; Chan, Jegadeesh, & Lakonishok, 1996; Chui, Titman, & Wei, 2010; Gutierrez & Kelley, 2008; Israel & Moskowitz, 2013; Jegadeesh & Titman, 1993; Korajczyk & Sadka, 2004; Rouwenhorst, 1998). Momentum investing has been documented to perform best in the short term (see, e.g., Cooper, Gutierrez, & Hameed, 2004; Jegadeesh & Titman, 2001; Lam, Liu, & Wong, 2010; Novy-Marx, 2012), whereas value investing performs better for longer holding periods (see, e.g., Bird & Casavecchia, 2007b; Lakonishok, Shleifer, & Vishny, 1994; Pätäri, Leivo, & Honkapuro, 2010). Since the prices of value stocks may remain low for an extended period, some scholars have begun to examine whether the performance of pure value investing strategies could be enhanced by complementing the portfolio-formation process with a timing indicator that shows when to purchase value stocks. To the best of our knowledge, the interaction of value and momentum strategies was first discussed by Asness (1997), who concluded that momentum and value are negatively correlated across stocks, yet each is positively related to the cross-section of

* Corresponding author.
E-mail addresses: eero.patari@lut.fi, eero.patari@gmail.com (E. Pätäri).

[1] By definition, value stocks have high fundamentals-to-price ratios and/or high fundamentals-to-enterprise value ratios, whereas the corresponding ratios for the glamour stocks are low. In this study, we use *book-to-price* (B/P), four variants of *cash flow-to-price* (CF1/P, CF2/P, CF3/P, and CFO/P), *dividend-to-price* (D/P), *earnings-to-price* (E/P), *sales-to-price* (S/P), *earnings before interest and taxes-to-enterprise value* (EBIT/EV), *earnings before interest, taxes, depreciation, and amortization-to-enterprise value* (EBITDA/EV), *free cash flow-to-enterprise value* (FCF/EV), and *sales-to-enterprise value* (S/EV) as valuation ratios.

http://dx.doi.org/10.1016/j.ejor.2017.08.001
0377-2217/

average stock returns. Recently, Barroso and Santa-Clara (2015) reported a highly significant negative relation between momentum- and B/P-based long-short portfolios. Some empirical evidence on the added value of combining value and momentum strategies has also been documented (see, e.g., Asness, Moskowitz, & Pedersen, 2013; Asness, Ilmanen, Israel, & Moskowitz, 2015; Bird & Casavecchia, 2007a; Cakici & Tan, 2014; Fisher, Shah, & Titman, 2016). However, in most studies, the selection of value-momentum portfolios has been based on double-sorting the stocks according to both of these two style dimensions, instead of merging both dimensions into a single efficiency score (to the best of our knowledge, the only exception is Pätäri, Leivo, & Honkapuro, 2012, who use several variants of data envelopment analysis to create such an efficiency score).

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 value- and momentum-based sub-criteria than in any of the earlier studies. Our analysis is based on 12 different individual valuation ratios and 6-month price momentum to examine the potential benefits of combining value and momentum criteria. Four of these 12 valuation ratios are enterprise value (EV) multiples, three of which have not been examined before in this context,[2] although the results of the few related value-only studies have been encouraging (see, e.g., Gray & Vogel, 2012; Loughran & Wellman, 2011; Walkshäusl & Lobe, 2015). 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 (McLean & Pontiff, 2016, show a remarkable post-publication decline in anomalous returns. See also the theory of gradual decay of anomalies introduced by Jones & Pomorski, 2017 and Pätäri & Leivo, 2017, for a comprehensive literature review on value anomalies). In addition, leverage is only implicitly included in price-based multiples, whereas it has direct influence on EV multiples.[3] We combine these 13 criteria by employing several multicriteria decision-making (MCDM) methods,[4] including simple median-scaling (MS), the additive DEA (add.DEA), the Analytic Hierarchy Process (AHP), and the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to find out whether the performance of value, momentum, or combined value-momentum portfolios and/or the value premium, winner-loser premium, or the top-bottom decile return spread stemming from combining value and momentum indicators can be enhanced from that generated by the best single selection criteria. As far as we know, Pätäri et al. (2010) were the first to document the applicability of DEA to detecting undervalued stocks by capturing several dimensions of relative value simultaneously. In their follow-up research, the same authors extended the variable combinations to include a price momentum indicator, which was the first instance of indicators of relative value and price momentum being combined into a single selection criterion (Pätäri et al., 2012). Both of these papers focused on MCDM methods based exclusively on DEA. Moreover, companies with negative earnings figures were always ranked in the bottom-tertile portfolios because the DEA applications employed in those studies could not cope with DEA variables that take positive values for some companies and negative values for others. In this paper, we overcome such a limitation by using scores based on the add.DEA model, AHP, and TOPSIS, which, as far as we know, have yet to be employed for this purpose in the previous literature.[5] This extension is important because there are many turnaround cases among negative income companies whose stocks are classified as value stocks according to criteria other than earnings or cash flow multiples (As an evidence of this, Penman & Reggiani, 2013 document high average book-to-price (B/P 0.98) for the lowest earnings yield (E/P) decile portfolio, which, on average, consisted of negative earnings stocks over the 1963–2006 sample period).

We evaluate the relative performance of portfolios formed on each selection criterion from several different viewpoints. We use raw returns, Sharpe ratios, and multi-factor alphas for this purpose. We calculate multi-factor alphas based on the Carhart (1997) 4-factor model to find out whether the best-performing portfolios based on each of the four MCDM methods have different exposures to the four most commonly used explanatory factors: market excess return (over and above risk-free return), size, value, and momentum factors. We employ comprehensive U.S. sample data that enables us to test the performance of combination strategies in the equity market, where pricing efficiency is generally assumed to be the highest in the world. We also repeat all the main sample tests for the large-cap sample consisting of 40% of the largest-cap U.S. stocks to see whether our results are robust to size effect, which refers to the tendency of small-cap stocks to outperform large-cap stocks in the long term and on average. 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.

The rest of the paper is organized as follows. Section 2 describes the data and the variables, while Section 3 explains the methodology employed. Section 4 introduces the empirical results supplemented by appropriate robustness tests. Section 5 summarizes the main findings and also discusses the limitations of the study, while Section 6 concludes.

2. Data

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 databases. 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, McNichols, and Price (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 1971–2013 period (We used estimates only in 0.65% of cases, which implies that our sample is essentially free of delisting bias).

[2] The only exception is EBITDA/EV that was included in Pätäri et al. (2012). In the comparison of three different DEA variants, the authors reported the best portfolio performance for the combination criteria that included EBITDA/EV as one of sub-criteria.

[3] E.g., Pätäri, Karell, and Luukka (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 and Leibowitz, 2002, for a comprehensive analysis of the relationship between E/P ratios and leverage).

[4] We refer to all these methods as MCDM methods, although a stricter definition for MCDM methods separates DEA from this category of methods (e.g., see Doyle and Green, 1993; Stewart, 1994, 1996; Bouyssou, 1999, for a discussion on linkages and differences between DEA and MCDM. By contrast, in a MCDM literature review of Ho, Xu, and Dey, 2010, DEA is classified as one category of MCDM methods). According to the same definition, median-scaling would neither be considered as a MCDM method.

[5] Although a few studies have employed one or more of these MCDM methods for equity portfolio selection (see, e.g., Saaty, Rogers, & Pell, 1980; Tiryaki & Ahlatcioglu, 2005; Hwang, Lin, & Chuang, 2007; Sevastjanov & Dymova, 2009; Ho & Oh, 2010), none of them has focused on the variables that specifically represent measures of relative value and momentum.

We use 13 different selection criteria to rank the stocks into decile portfolios. Twelve out of 13 criteria are value measures, of which eight are price-based ratios (i.e., B/P, CF1/P, CF2/P, CF3/P, CFO/P, D/P, E/P, and S/P) and four are EV-based ratios (i.e., EBIT/EV, EBITDA/EV, FCF/EV, and S/EV). Motivated by recent evidence regarding added value through the combination of value and momentum indicators (see, e.g., Asness et al., 2013, 2015; Bird & Casavecchia, 2007a; Leivo & Pätäri, 2011), we also include one price momentum indicator that ranks the stocks based on 6-month historical returns with one month lag for a 12-month holding period.[6] The components of the valuation ratios are calculated according to the following principles[7]:

- Market Value of Equity (ME): ME is the stock price multiplied by shares outstanding from the CRSP monthly file and obtained at the 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): Like 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, Xue, and Zhang (2015), we compute CF1 as Income Before Extraordinary Items (item IB) plus Equity's Share of Depreciation plus Income Statement Deferred Taxes (item TXDI). For Equity's Share of Depreciation, we adopt Fama and French's (2006a) definition: ME/[Total Assets (item AT) – BE (item SEQ) + ME]*Depreciation and Amortization (item DP).
- Cash flow 2 (CF2): Similar to Dhatt, Kim, and Mukherji (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, Karolyi, and Kho (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 Desai, Rajgopal, and Venkatachalam (2004), we compute CFO as Earnings (E) minus Accruals. Accruals are computed according to Sloan's (1996) definition, 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.
- Free Cash Flow (FCF): Following 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 our final sample, the firms must have all the information available for the calculation of each of the 13 single selection criteria being examined at each portfolio-formation point. Although this prerequisite reduced the number of otherwise usable firm-year observations, we require this information in order to ensure the best possible comparability of the results based on different single selection criteria and/or combination criteria. For each valuation ratio, we checked case-by-case those values that were outside the range of 2.5 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 errors (the most common reason for this was the omission of either the reverse split in cases of extremely high returns or split in cases of extremely low returns). To alleviate backfill bias, in accordance with Fama and French (1993), firms are required to have two-year 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. In addition, only firms with fiscal year durations of 12 months are included in the final sample. Consistent with the existing literature (see, e.g., Fama & French, 2008; Loughran & Wellman, 2011), we also exclude firm-year observations for which the book value of equity is negative. The same practice is also followed in cases of negative EV. As in Jegadeesh and Titman (2001), and Avramov, Chordia, Jostova, and Philipov (2007), the firms for which market capitalization at the beginning of each 1-year holding period is below the bottom NYSE decile breakpoint are excluded, as well as those that had traded below $5 at each checkpoint. After all exclusion criteria, our final sample contains 48,390 firm-year observations with complete data for each of the 13 selection criteria over the 1971–2013 sample period (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).

3. Methodology

The equal-weighted decile portfolios are formed at the end of April of year *t* and held for one year from May of year *t* through

[6] We use this particular combination of selection and holding period because we would only like to reform the portfolios once a year, consistent with their value-only counterparts, and because based on previous literature, the optimal length of the selection period in such cases would be approximately six months (see, e.g., Scowcroft & Sefton, 2005; Figelman, 2007). The 6-month price momentum has also been documented to work in earlier studies in which quantile portfolios have been formed on both value and price momentum indicators (see, e.g., Bird & Casavecchia, 2007a, 2007b; Leivo & Pätäri, 2011).

[7] The same principles are also followed in the calculation of input and output variables for the add.DEA method, with the exception that in such cases, variables are converted to a per-share form. In the case of the add.DEA, the output variable representing the momentum indicator is defined by multiplying the total return of a stock over the period from the beginning of October *t* – 1 to the end of March *t* by the corresponding input variable, consistent with Pätäri et al. (2012).

April of year $t+1$. Like Fama and French (1992, 1993) and Davis, Fama, and French (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 statements after the end of the fiscal year (see Fama & French, 1992). By contrast, the market values of equity used in denominators of price multiples and EV multiples are updated to match those prevailing at the end of April in year t for all firms, consistent with Desai et al. (2004), Lakonishok et al. (1994), and Fisher et al. (2016), for example. Six-month momentum indicators are calculated by skipping the return for the most recent month, as is the convention in the momentum literature (see, e.g., Asness et al., 2013; Barroso & Santa-Clara, 2015; Fama & French, 2008, 2012; Israel & Moskowitz, 2013; Novy-Marx, 2012). The stocks with either the highest valuation ratios, 6-month historical returns, or the highest efficiency scores based on each combination criteria are placed in the top decile, whereas the bottom deciles consist of the stocks with the lowest valuation ratios, momentum, or combined efficiency scores. 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.[8] As a result, we get monthly returns for each decile portfolio over a 42-year investment period.

3.1. MCDM methods used for portfolio allocation

To generate the efficiency scores for each stock at each checkpoint, we use four different combination methods that are median-scaling, the TOPSIS (developed by Hwang & Yoon, 1981), the AHP (Saaty, 1977), and the add.DEA (Charnes, Cooper, Golany, Seiford, & Stutz, 1985). We focus on these four methods because they are relatively well-known and widely-used. They also represent different methodological approaches and they can deal appropriately with negative data, either as such or with minor adjustments. Analogous to Dhatt et al. (2004), the median-scaled composite measures are calculated by first standardizing each metric of a single selection criterion of a firm in a particular year by its cross-sectional median at the same time point and then computing the average of these median-scaled scores for each combination.[9] For example, if the B/P for a firm at some point is 0.9, while 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. The corresponding scores for momentum are calculated by dividing the past 6-month return by the median 6-month sample return at the same time point (both returns expressed in an investment relative form, i.e., as $1+r$). Although plenty of variants for the basic TOPSIS, AHP, and add.DEA methods have been suggested since their introduction,[10] we employ their simple variants, because our focus is on the comparison of different methodological approaches, and not on the comparison of methodological variants within each of the methods (This limitation is also necessary for space limitations). In addition, as shown by abundance of related literature, every variant has pros and cons of its own, and none of them is superior to the others in all cases (for the overview of developments in the AHP, TOPSIS, and DEA methods, see Behzadian, Otaghsara, Yazdani, & Ignatius, 2012; Ishizaka & Labib, 2011; Cook & Seiford, 2009, respectively). Therefore, the added-value of some methodological refinement may, depending on the case, as well be negative, as it may be positive. One additional justification for our research design is that these basic methodologies being compared, unlike some of their extensions, are simple enough for being adopted by equity investors (e.g., according to Zopounidis, Galariotis, Doumpos, Sarri, & Andriosopoulos, 2015, sophisticated analytic methodologies are not often used in practice, because they may often be far too complex for financial decision-makers to understand). In addition, both the TOPSIS and AHP can easily be modified to eliminate all the subjectivity from the selection process, the removal of which is desirable in our research design. In case of the add.DEA, we use of a slacks-based measure (SBM) model introduced by Tone (2001), which already eliminates this by also providing efficiency scores for all inefficient units.

To avoid ranking paradoxes in the AHP stemming from negative values, we first scale all the single criteria values at each portfolio-formation point to positive values with the upper limit in unity. Next, pairwise comparisons within each of the single criteria are done by calculating the ratios of these scaled values for each possible pair of the two alternatives x_h/x_k, where x_h and x_k represent either the scaled valuation ratios or the corresponding momentum indicator for the companies h and k at each portfolio-formation point. Otherwise, the AHP procedure used in this study is similar to Saaty (1977). In the calculation of efficiency scores based on the TOPSIS, each sub-criterion is given equal weight when determining the weighted normalized decision matrix in step 2 (e.g., see Behzadian et al., 2012, for stepwise description of the TOPSIS procedure) because we do not want to make any subjective presumptions on the relative importance of some sub-criterion (or sub-criteria) over the others.

3.2. Test procedures for performance comparisons

We test a large number of combination strategies by forming portfolios on all possible 2-, 3-, and 4-combinations that are based on the four MCDM criteria introduced above (We also tested 5- and 6-combinations but their marginal utility compared to 4-combinations was mostly negative; therefore, we restrict our analysis to 2-, 3-, and 4-combinations). The 2-combinations based on median-scaling, TOPSIS, and AHP are formed by combining either two individual valuation ratios or one individual valuation ratio and the price momentum indicator. The corresponding 3-combinations (4-combinations) are formed on three (four) single criteria, respectively. In the context of the add.DEA, 2-combinations are defined as those including two output variables and one input variable, whereas the corresponding 3-combinations include three output variables and one input variable. DEA 4-combinations include either four output variables and one input variable or two output variables and two input variables (the latter type of 4-combinations implicitly includes four output/input ratios since DEA is based on all possible output/input combinations). As valuation ratios are actually output/input ratios, we use 12 different output variables (earnings, EBIT, EBITDA, CF1, CF2, CF3, CFO, FCF, book equity, dividends, sales, and momentum indicator) and two input variables (i.e., market value of equity and EV) as potential constituents of DEA efficiency scores.

The performance of decile portfolios is primarily evaluated based on the average return and the Sharpe ratio (Sharpe, 1966). In order to avoid validity problems stemming from the negative excess returns in the context of the Sharpe ratio comparisons, we use a refined version of the Sharpe ratio introduced by

[8] 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 both value and momentum anomalies (e.g., see Loughran & Wellman, 2011; Korajczyk & Sadka, 2004), 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 the pros and cons of the use of value-weighted returns in anomaly studies.

[9] A prerequisite for the validity of median-scaling is that medians are all positive, as is the case for all the ratios employed in this study.

[10] E.g., see Chang, Lin, Lin, and Chiang (2010), Chen Li, and Liu (2011), Dymova, Sevastjanov, and Tikhonenko (2013), Kuo (2017), for the TOPSIS variants, and Lin (2007), Saaty and Shang (2011), Kou and Lin (2014), Zhu, Xu, Zhang, and Hong (2016), for the AHP variants, and Du, Chen, Chen, Cook, and Zhu (2012), Banker, Chen, and Klumpes (2016), and Sahoo, Khoveyni, Eslami, and Chaudhury (2016), for the add.DEA variants.

Israelsen (2005), as follows:

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

where r_i is the average monthly return of a portfolio i, r_f is the average monthly risk-free rate of the return, σ_i is the standard deviation of the monthly excess returns of a portfolio i, and ER is the average excess return of portfolio i.

To evaluate whether the potential abnormal returns are explained by four commonly used explanatory factors, we also calculated the Carhart (1997) 4-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. The regression equation is as follows (all the factor returns were downloaded from French's 2015 data library, available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html):

$$r_{it} - r_{ft} = \alpha_i + b_i\left(r_{mt} - r_{ft}\right) + s_i SMB_t + h_i HML_t + m_i WML_t + \varepsilon_{it} \tag{2}$$

where r_{it} is the return of a portfolio, r_{ft} is the risk-free rate of return, α_i is the 4-factor alpha (the abnormal return over and above what might be expected based on the 4-factor model employed), r_{mt} is the stock market return,[11] SMB_t is the return of the size factor (i.e., the return difference between small- and large-cap portfolios), HML_t is the return of the book-to-market (B/P) factor (i.e., the return difference between high- and low-B/P portfolios), WML_t is the return of the momentum factor (i.e., the return difference between winner and loser stock portfolios), b_i, s_i, h_i, and m_i are factor sensitivities to stock market, SMB, HML, and WML factors, respectively, and ε_{it} is the residual term.

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.[12] We also test the significance of 4-factor alphas based on their t-statistic, which is closely related to the Treynor and Black (1973) multifactor appraisal ratio, also known as the information ratio (see, e.g., Lewellen, 2010). In addition, we test the statistical significance for the difference between the top- and bottom-decile portfolio alphas using the appropriate alpha spread test, following Kosowski, Naik, and Teo (2007). We also test monotonicity in the returns, as well as in the alphas of decile portfolios formed on each portfolio-formation criterion by implementing the monotonic relation test (MR-test) developed by Patton and Timmermann (2010).[13] As a supplementary test, we also calculate the Gibbons, Ross, and Shanken (1989) test statistic to find out whether the null hypothesis of jointly zero decile alphas holds for each portfolio-formation criterion (Unlike the MR test statistic, the GRS statistic is not affected by the order of alphas. On the other hand, the former is based on the differences between the decile alphas, and not affected by overall level of alphas, unlike the latter that uses zero as the benchmark). Throughout the study, we use Newey and West (1987) standard errors in the statistical tests in order to avoid problems related to autocorrelation and heteroscedasticity. Moreover, we carried out Jarque and Bera's (1980) normality test for regression residuals, but the normality assumption was never violated. We also tested for the existence of multicollinearity in our multifactor regression models and found that the variance inflation factors were low in all cases.

4. Results

4.1. The average results for the combination methods

The annualized geometric average returns for all combination methods employed are shown in Table 1, where Panels A, B, and C indicate the corresponding average decile returns for all 2-, 3-, and 4-combinations (Henceforth, all the reported returns are annualized geometric returns, unless otherwise indicated). Because some valuation ratios are close proxies for each other, we restrict the number of potential combinations so that only one of earnings or cash flow variables included in price-based multiples (i.e., earnings, CF1, CF2, CF3, or CFO) and one of the corresponding variables included in EV multiples (i.e., EBIT, EBITDA, or FCF) at a time can occur in the combinations when calculating the averages for 2-, 3-, and 4-combinations reported in Table 1. This implies that the reported averages are based on 65 2-combinations, 165 3-combinations, and 235 4-combinations in cases of median-scaling, the AHP, and TOPSIS. The corresponding number in cases of the add.DEA is 53 for 2-combinations (for both input variables) and 112 for 3-combinations, respectively. The number of corresponding 4-combinations with only one input variable is 123 in both cases, whereas with two inputs it is 65.

Table 1 indicates that the average returns of top-decile portfolios, as well as the corresponding top-bottom decile return spreads, generally increase monotonically when moving from 2-combinations to 3-combinations, and likewise when moving from 3-combinations to 4-combinations for all the combination methods other than median-scaling (MS), which shows exactly the opposite pattern. The highest average return (17.46%) among the decile portfolios is documented for the top-decile TOPSIS 4-combinations, whereas the highest average top-bottom decile return spread (9.63%) is generated by AHP 4-combinations (the corresponding equal-weighted sample return is 13.04% p.a. for the same period).

Table 2 shows the corresponding statistics in terms of total risk-adjusted returns. The results remain very similar to those based on raw returns: the average annualized Sharpe ratio is the highest for the top-decile TOPSIS 4-combinations (0.608), followed by the corresponding AHP portfolios (0.603), whereas for the equal-weighted benchmark portfolio that consists of the sample stocks, it is 0.369. Analogous to the average combination returns, higher level combinations tend to dominate lower level combinations in terms of the average Sharpe ratios for MCDM methods other than that based on median-scaling. Thus, on average, the investor would

[11] 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 left-hand side portfolios are equal-weighted. This also makes the 4-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 main 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 small-cap anomaly (for recent evidence of this, see e.g., Israel & 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 was only 6.9% p.a. during the sample period).

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

[13] The corresponding programming code is freely available on Patton's home page at: http://public.econ.duke.edu/~ap172/code.html.

Table 1
Average decile returns for the combination methods.

Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: averages for 2-combinations											
TOPSIS	17.03	15.92	15.29	14.51	13.48	12.46	12.00	11.07	9.24	9.40	7.63
AHP	16.90	15.90	15.17	14.46	13.58	12.69	11.98	11.06	9.60	8.96	7.94
MS	16.20	15.19	14.77	14.13	13.49	12.76	12.04	11.69	10.34	9.93	6.27
add.DEA (i: EV)	16.05	13.52	12.83	13.37	13.56	13.08	12.76	12.74	11.82	10.65	5.40
add.DEA (i: ME)	16.00	13.63	14.65	13.20	13.44	13.03	12.04	12.46	11.50	10.42	5.57
Panel B: averages for 3-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	17.23	16.00	15.26	14.57	13.59	12.48	11.97	11.10	9.14	9.03	8.20
AHP	17.03	15.84	15.07	14.44	13.98	13.20	12.03	10.83	9.82	8.14	8.89
add.DEA (i: EV)	16.34	13.73	12.85	13.18	13.71	12.98	12.64	12.67	11.74	10.53	5.81
add.DEA (i: ME)	16.23	13.93	14.69	13.18	13.41	13.02	11.96	12.34	11.41	10.2	6.03
MS	15.89	15.09	14.79	13.81	13.34	12.93	11.95	12.08	10.66	9.84	6.06
Panel C: averages for 4-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	17.46	15.99	15.22	14.48	13.75	12.40	12.03	11.12	9.16	8.76	8.70
AHP	17.18	15.74	15.20	14.30	14.14	13.14	12.39	10.78	9.96	7.55	9.63
add.DEA (i: EV)	16.74	13.92	13.01	13.00	13.92	12.84	12.54	12.39	11.59	10.42	6.31
add.DEA (i: ME)	16.53	14.31	14.52	13.41	13.25	13.01	11.99	12.09	11.24	10.03	6.50
add.DEA (i: EV, ME)	15.96	14.17	13.56	13.48	13.40	12.98	12.42	12.42	11.57	10.37	5.59
MS	15.59	14.89	14.80	13.56	13.16	13.02	11.83	12.39	11.23	9.90	5.70

Note: The decile averages are reported separately for the add.DEA combinations that use market equity (ME), enterprise value (EV), or both (ME & EV) as input(s).

Table 2
Average Sharpe ratios for the combination methods.

Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
Panel A: averages for 2-combinations											
TOPSIS	0.579	0.549	0.519	0.475	0.412	0.346	0.306	0.243	0.151	0.144	0.435
AHP	0.570	0.529	0.502	0.463	0.412	0.354	0.306	0.248	0.171	0.135	0.435
MS	0.544	0.509	0.489	0.451	0.402	0.348	0.305	0.271	0.192	0.165	0.378
add.DEA (i: ME)	0.478	0.366	0.396	0.346	0.362	0.345	0.312	0.344	0.302	0.299	0.179
add.DEA (i: EV)	0.462	0.354	0.333	0.355	0.372	0.346	0.339	0.359	0.317	0.321	0.141
Panel B: averages for 3-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	0.594	0.556	0.522	0.482	0.415	0.347	0.307	0.240	0.145	0.132	0.461
AHP	0.588	0.531	0.495	0.460	0.420	0.367	0.305	0.242	0.175	0.114	0.474
MS	0.532	0.498	0.485	0.434	0.392	0.349	0.301	0.290	0.210	0.172	0.36
add.DEA (i: ME)	0.488	0.377	0.392	0.348	0.360	0.344	0.311	0.339	0.300	0.292	0.196
add.DEA (i: EV)	0.476	0.357	0.335	0.349	0.377	0.344	0.331	0.357	0.317	0.316	0.160
Panel C: averages for 4-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	0.608	0.557	0.522	0.484	0.422	0.343	0.309	0.238	0.145	0.123	0.485
AHP	0.603	0.537	0.499	0.452	0.424	0.365	0.312	0.240	0.178	0.098	0.505
MS	0.519	0.487	0.479	0.419	0.384	0.347	0.295	0.302	0.233	0.184	0.335
add.DEA (i: ME)	0.502	0.385	0.388	0.355	0.358	0.344	0.314	0.331	0.294	0.283	0.219
add.DEA (i: EV)	0.486	0.362	0.338	0.346	0.377	0.343	0.329	0.351	0.316	0.311	0.175
add.DEA (i: EV, ME)	0.466	0.382	0.357	0.359	0.360	0.345	0.325	0.348	0.309	0.305	0.161

Note: The decile averages are reported separately for the add.DEA combinations that use market equity (ME), enterprise value (EV), or both (ME & EV) as input(s).

benefit more from using operations research-based MCDM methods (TOPSIS, the AHP, and the add.DEA) for 4-combinations (3-combinations) than for 3-combinations (2-combinations) in terms of both raw returns and Sharpe ratios. However, it should be noted that the average decile returns reported for the add.DEA portfolios are not entirely comparable to those based on other MCDM methods examined, since the number of combinations included in the calculation of averages varies not only between the add.DEA and other methods (owing to methodological differences in forming the set of ratios included), but also within the add.DEA 4-combinations, depending on whether one or two input variables are employed (see the first paragraph of this section for details). Therefore, we next examine the best combination criteria within each of the four combination methods employed. These combination criteria are chosen on the basis of the highest sample-period returns and Sharpe ratios of the top-decile portfolios within each of the four combination methods, resulting in eight composite criteria that are henceforth called top combinations (These top-decile portfolios are also the best throughout all the decile portfolios).

Table 3
Decile portfolio performance for the top combinations and the best single selection criteria.

Panel A: raw returns											
Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS2 (EBITDA/EV, S/P)***	19.05	16.37	15.21	14.22	14.26	12.03	11.60	10.97	8.22	8.46	10.59
	(0.1)	(4.1)	(7.7)	(19.5)	(29.0)	(27.4)	(10.2)	*(1.6)*	*(0.0)*	*(0.2)*	(0.0)
MS4 (B/P, EBITDA/EV, MOM, S/EV)**	18.93	16.06	16.35	14.94	12.96	13.10	11.73	10.87	10.21	5.24	13.69
	(0.2)	(4.0)	(0.5)	(9.8)	(93.7)	(99.2)	(7.3)	*(0.8)*	*(1.1)*	*(0.0)*	(0.0)
AHP4 (CF3/P, FCF/EV, MOM, S/EV)*	18.54	15.20	14.63	15.11	14.43	13.64	13.84	11.96	8.91	4.13	14.40
	(0.0)	(26.9)	(26.2)	(4.3)	(10.7)	(40.7)	(26.8)	(29.1)	*(0.1)*	*(0.0)*	(0.0)
add.DEA3 (i: EV; o: BE, DIV, MOM)	18.09	14.34	13.32	11.79	13.02	13.45	12.39	11.85	11.69	10.43	7.66
	(1.4)	(92.1)	(56.4)	(1.6)	(31.9)	(98.2)	(11.2)	(30.3)	(56.0)	(52.8)	(0.3)
EBIT/EV	17.80	14.93	15.60	13.96	12.71	13.03	11.82	10.57	9.82	10.15	7.64
	(0.2)	(8.8)	(2.8)	(49.2)	(97.7)	(91.1)	(8.9)	*(0.5)*	*(0.3)*	*(1.0)*	(0.1)
Panel B: Sharpe ratios											
Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
AHP4 (CF3/P, D/P, MOM, S/EV)	0.688	0.514	0.547	0.398	0.394	0.353	0.321	0.312	0.230	-0.001	0.689
	(0.1)	(10.7)	(0.8)	(65.9)	(71.2)	(58.1)	(32.3)	(26.3)	*(2.2)*	*(0.0)*	(0.0)
TOPSIS2 (D/P, EBITDA/EV)***	0.686	0.553	0.553	0.515	0.402	0.360	0.238	0.236	0.144	0.118	0.568
	(0.0)	(6.4)	(3.1)	(2.7)	(86.5)	(42.4)	*(0.1)*	*(0.7)*	*(0.1)*	*(0.3)*	(0.0)
MS4 (CFO/P, D/P, MOM, S/EV)	0.651	0.511	0.495	0.433	0.314	0.342	0.265	0.297	0.258	0.148	0.503
	(0.5)	(26.2)	(18.5)	(44.2)	(18.0)	(62.4)	*(0.4)*	(17.5)	(7.7)	*(0.0)*	(0.0)
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)
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)*	0.541	0.364	0.374	0.331	0.350	0.354	0.301	0.309	0.295	0.279	0.262
	(0.2)	(99.0)	(90.3)	(43.5)	(40.4)	(49.1)	(13.0)	(15.1)	(21.6)	(34.3)	(0.1)

Note: Panels also indicate the significance levels (in percentages in parentheses) for the Sharpe ratio difference between each decile portfolio and the market portfolio. The last column shows the corresponding top-bottom decile (D1–D10) differences, respectively (All these significance levels are based on the Ledoit–Wolf test and italicized in case of significant underperformance). The significances for the MR test for decile returns are shown by asterisks. * (**/***) indicates that decile returns are monotonically decreasing at the 1% (5%/10%) significance level. MS refers to median-scaling, i: (o:) refers to input (output) variable(s) included in the add.DEA method, ME to market equity, BE to book equity, DIV to dividends, and MOM to momentum.

4.2. The results for the top combination criteria

Table 3 Panel A shows the results for the top combinations in terms of raw returns. Each combination method is capable of generating higher top-decile portfolio returns than the best single selection criterion (i.e., EBIT/EV, 17.80%).[14] The highest return (19.05%) is documented for the top-decile TOPSIS2 (EBITDA/EV, S/P) portfolio, whereas the highest top-bottom decile return spread (14.40%) is generated by the AHP4 (CF3/P, FCF/EV, MOM, S/EV) criterion. Generally, the top combinations can generate much larger top-bottom decile return spreads than the single selection criteria, implying that these combinations can be useful not only for the long-only portfolio selection, but also for forming long-short portfolios (among the 13 single selection criteria, clearly the highest spread (11.48%) between the top- and bottom-decile returns is generated for the momentum criterion, mainly owing to a low return (5.82%) of the loser decile portfolio (i.e., the bottom portfolio)). However, the top combinations that generate the highest top-decile returns within each combination method are never based on exactly the same variables. The most frequently appearing sub-criterion within these top combinations is momentum, which is included in three out of four cases. This finding is not surprising, as momentum is the only representative of its own style dimension, whereas value dimension has 12 different proxies in our research design.

The total risk-adjusted return comparison of the top combinations reveals that D/P, which generates the highest Sharpe ratio among the single selection criteria (0.621), is included in all four cases (Table 3, Panel B). Hence, the low-risk characteristic of high dividend yield stocks seems to also extend to the top-decile portfolios formed on those combination criteria in which D/P is a sub-criterion.[15] Compared to the top combinations that generate the highest top-decile raw returns (in Panel A), none of their total risk-adjusted counterparts (in Panel B) is exactly the same. Interestingly, the two top criteria based on AHP are 4-combinations, whereas based on TOPSIS, they are 2-combinations, although, on average, TOPSIS-based 4-combinations have performed somewhat better than the corresponding 2-combinations, as have also done their AHP- and add.DEA-based counterparts (see Tables 1 and 2).

[14] For the same sample over the same period, the corresponding detailed results for the same 12 valuation ratios as included in our analysis are reported in Pätäri et al. (2017).

[15] This inference is reinforced by the performance decomposition based on bullish and bearish months reported in Section 4.3. Pätäri et al. (2017) show that among the individual valuation ratios, the total risk-adjustment (quantified by the Sharpe ratio) has particularly strong impact on the top-decile ranking of the corresponding D/P portfolio, which has remarkably lower volatility than the corresponding top-decile portfolios chosen on the basis of any other valuation ratio (while the top-decile D/P portfolio generated the highest Sharpe ratio, its raw return was the lowest (15.04% p.a.) among the top-decile portfolios formed on the 12 valuation ratios). 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% p.a.) for the D/P-based top decile. 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. 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.

Table 4
4-factor decile alphas for the top combination and the best single selection criteria.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10	GRS
add.DEA3 (i: EV; o: BE, DIV, MOM)[1]	4.62	0.42	−0.31	−1.42	−0.74	0.22	−1.19	−1.19	−0.71	−0.69	5.31	3.05
	(0.1)	(62.2)	(67.1)	*(4.4)*	(28.1)	(75.1)	(9.2)	(10.6)	(26.7)	(24.5)	(0.0)	(0.1)
TOPSIS2 (EBITDA/EV, S/P)[1] ***	3.72	1.70	0.71	0.07	−0.07	−1.29	−1.21	−0.55	−1.40	−0.45	4.17	4.01
	(0.0)	(2.1)	(34.8)	(93.2)	(92.3)	(6.3)	(8.5)	(47.4)	(22.1)	(78.4)	(2.5)	(0.0)
AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1] ***	3.69	0.67	0.08	0.92	0.38	−0.34	0.43	−0.45	−1.93	−4.43	8.12	2.37
	(0.0)	(45.5)	(91.2)	(15.2)	(53.0)	(59.9)	(55.4)	(52.0)	*(4.3)*	*(0.4)*	(0.0)	(1.0)
MS4 (B/P, EBITDA/EV, MOM, S/EV)[1]	3.53	1.18	1.69	0.40	−1.19	−0.44	−1.26	−1.08	−0.19	−2.49	6.02	3.21
	(0.0)	(8.5)	(1.4)	(59.6)	(6.0)	(52.6)	*(4.2)*	(9.3)	(83.0)	(8.7)	(0.1)	(0.1)
EBIT/EV[1]	3.46	1.01	1.43	−0.13	−1.17	−0.55	−1.04	−0.96	−0.92	−0.46	3.93	4.33
	(0.0)	(18.8)	(4.7)	(86.9)	(9.7)	(44.1)	(12.1)	(23.3)	(47.9)	(78.8)	(4.0)	(0.0)
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2] **	3.05	0.21	−0.13	−0.69	−0.50	−0.28	−0.90	−1.01	−0.60	−0.32	3.37	1.46
	(1.3)	(80.7)	(86.8)	(32.2)	(47.0)	(68.7)	(24.8)	(17.2)	(39.7)	(58.8)	(1.3)	(15.1)
TOPSIS2 (D/P, EBITDA/EV)[2]	2.85	1.46	1.10	1.18	−0.70	−0.67	−2.03	−0.53	−0.61	−0.51	3.36	5.30
	(0.1)	(6.3)	(14.2)	(8.5)	(36.0)	(33.5)	*(0.6)*	(52.6)	(57.4)	(75.0)	(6.5)	(0.0)
AHP4 (CF3/P, D/P, MOM, S/EV)[2]	2.47	0.40	1.48	−0.64	−0.20	−0.05	−0.02	0.36	0.43	−4.04	6.51	3.03
	(0.1)	(60.2)	(4.7)	(36.6)	(76.8)	(94.5)	(97.4)	(63.3)	(66.2)	*(0.3)*	(0.0)	(0.1)
MS4 (CFO/P, D/P, MOM, S/EV)[2]	1.95	0.27	−0.23	−0.04	0.09	1.16	−0.84	0.32	0.10	−2.29	4.24	1.90
	(4.5)	(77.3)	(77.8)	(95.8)	(91.5)	(16.7)	(26.6)	(69.4)	(92.6)	(8.1)	(0.9)	(4.3)
D/P[2]	1.03	0.01	0.17	0.73	−0.71	−1.65	0.07	−1.67	−1.02	−1.94	2.97	1.27
	(20.3)	(99.5)	(85.7)	(45.4)	(45.1)	(9.8)	(94.1)	(8.4)	(30.7)	(7.0)	(2.7)	(24.2)

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within either each of the four combination methods or among the 13 single selection criteria. Significances (in percentages) for the alphas and for the top-bottom decile alpha spreads (D1–D10) are in parentheses and italicized in case of significant underperformance. The last column shows the GRS statistics (significance levels in percentages in parentheses). The significances for the MR test for decile alphas are shown by asterisks. ** (***) indicates that decile alphas are monotonically decreasing at the 5% (10%) significance level.

The significance levels for the relative performance against the market portfolio reveal that the TOPSIS and AHP methods have the best discriminatory power in separating the outperforming stocks of the future from their underperforming counterparts. The highest Sharpe ratio (0.688) is documented for the top-decile AHP4 (CF3/P, D/P, MOM, S/EV) portfolio. Based on the Ledoit–Wolf test, the same combination criterion also generates the most significant performance difference between the top- and bottom-decile portfolios. By contrast, the most significant outperformance over the market portfolio (based on the same test) is documented for the top-decile portfolio formed on TOPSIS2 (D/P, EBITDA/EV). However, based on the MR test for raw returns, the highest discriminatory power is reported for add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM), for which the returns decrease monotonically throughout the deciles at the 0.1% significance level. In addition, the decreasing monotonicity in decile returns is also significant (at the 1% level) for AHP4 (CF3/P, FCF/EV, MOM, S/EV), and (at the 5% level) for MS4 (B/P, EBITDA/EV, MOM, S/EV), while being weakly significant (at the 10% level) for both TOPSIS combinations, and insignificant for the rest three. Although the results on the relative discriminatory power of the methods seem somewhat mixed, it is noteworthy that the relative performance may look different depending on whether it is evaluated on the basis of relative performance of the top-decile portfolio against the market portfolio, or based on the difference between the top- and bottom-decile portfolios, or in terms of differences between the decile portfolios.

Based on performance differences between the top- and bottom-decile portfolios, the add.DEA seems to have somewhat lower discriminatory power than the AHP and TOPSIS, in terms of both raw and total risk-adjusted returns. Also, the average results on the same performance metrics shown in Tables 1 and 2 give some support to this conclusion. However, it should be noted that different performance metrics may produce different results. Therefore, as a robustness check for performance comparisons, we next calculate the 4-factor alphas for all the decile portfolios formed on the eight top combinations and the best single selection criteria shown in Table 3. After the Sharpe ratio, the 4-factor alpha calculated on the basis of the Carhart (1997) factor model is the second most popular performance metrics employed in current financial literature and provides an alternative, factor-based approach to risk-adjustment.

4.3. Robustness checks

Among the decile portfolios formed on the eight top combination criteria, the highest 4-factor alpha (4.62% p.a.) is reported for the top-decile add.DEA3 (i: EV; o: BE, DIV, MOM) portfolio (Table 4). However, this combination is neither the best nor even among the best in performance comparisons that are based on total risk-adjusted returns, indicating that (multi)factor-based alphas and the total risk-based performance metrics represent different dimensions of performance. This result also shows that cautious is in place when comparing the general discriminatory power of different MCDM methods. In addition, it should be noted that the multifactor alphas were only calculated for the top combination strategies (based on the performance of top-decile portfolios) within each combination method. If they had been calculated for all possible combination portfolios, it would have been possible that some of them would have generated even higher multifactor alphas than this particular add.DEA combination. However, 4-factor alphas are for the most part in line with the other performance metrics as they are, on average, higher for the top-decile portfolios formed on the top combinations (3.24% p.a.) than they are for the corresponding portfolios formed on the 13 single selection criteria (2.26% p.a.). The same also holds for the corresponding top-bottom alpha spreads, of which the highest, as well as the most significant (8.12% with *t*-statistic of 4.49), is documented for AHP4 (CF3/P, FCF/EV, MOM, S/EV), indicating that this particular criterion would have worked well as the basis of a long-short strategy not only in terms of raw returns, but also in terms of the 4-factor alphas. Interestingly, all four combination criteria that belong to the top-8 combinations based on their highest top-decile return within each combination method, generate higher 4-factor alphas than any of the four other top combination criteria that are in the top-8 based on their highest Sharpe ratio within each combination method. In addition, within the top-4 and bottom-4 top-decile alpha rankings for the combinations, the rank order of the combination methods is the same: The add.DEA is followed by TOPSIS that is followed by

Table 5
4-factor regression results for the top- and bottom-decile portfolios of the top combination and the best single selection criteria.

Panel A: 4-factor regression results for top-decile portfolios

Portfolio-formation criterion	Alpha (%)	*p* (%)	MKT	*p* (%)	SMB	*p* (%)	HML	*p* (%)	WML	*p* (%)	*R*-squared
add.DEA3 (i: EV; o: BE, DIV, MOM)[1]	4.62	(0.1)	0.950	(0.0)	0.389	(0.0)	−0.214	(0.0)	0.089	(5.7)	0.916
TOPSIS2 (EBITDA/EV, S/P)[1]	3.72	(0.0)	0.991	(0.0)	0.095	(2.7)	0.468	(0.0)	−0.106	(0.0)	0.944
AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1]	3.69	(0.0)	0.985	(0.0)	0.145	(0.3)	0.252	(0.0)	−0.037	(22.8)	0.926
MS4 (B/P, EBITDA/EV, MOM, S/EV)[1]	3.53	(0.0)	0.982	(0.0)	0.066	(21.1)	0.498	(0.0)	−0.100	(0.0)	0.932
EBIT/EV[1]	3.46	(0.0)	0.941	(0.0)	0.118	(0.3)	0.338	(0.0)	−0.103	(0.0)	0.913
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2]	3.05	(1.3)	0.953	(0.0)	0.281	(0.0)	0.066	(14.5)	0.048	(23.2)	0.926
TOPSIS2 (D/P, EBITDA/EV)[2]	2.85	(0.1)	0.856	(0.0)	−0.037	(45.7)	0.502	(0.0)	−0.061	(1.1)	0.921
AHP4 (CF3/P, D/P, MOM, S/EV)[2]	2.47	(0.1)	0.842	(0.0)	−0.088	(13.3)	0.512	(0.0)	−0.034	(10.5)	0.914
MS4 (CFO/P, D/P, MOM, S/EV)[2]	1.95	(4.5)	0.847	(0.0)	−0.395	(0.0)	0.400	(0.0)	−0.011	(74.4)	0.870
D/P[2]	1.03	(20.3)	0.775	(0.0)	−0.163	(0.4)	0.516	(0.0)	−0.032	(12.5)	0.884

Panel B: 4-factor regression results for bottom-decile portfolios

Portfolio-formation criterion	Alpha (%)	*p* (%)	MKT	*p* (%)	SMB	*p* (%)	HML	*p* (%)	WML	*p* (%)	*R*-squared
add.DEA3 (i: EV; o: BE, DIV, MOM)[1]	−0.69	(24.5)	0.924	(0.0)	−0.716	(0.0)	−0.006	(84.9)	−0.012	(48.1)	0.942
TOPSIS2 (EBITDA/EV, S/P)[1]	−0.45	(78.4)	1.118	(0.0)	0.277	(4.3)	−0.854	(0.0)	0.077	(18.3)	0.882
AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1]	−4.43	(0.4)	1.081	(0.0)	0.155	(24.0)	−0.584	(0.0)	−0.096	(11.6)	0.880
MS4 (B/P, EBITDA/EV, MOM, S/EV)[1]	−2.49	(8.7)	1.071	(0.0)	0.144	(28.3)	−0.879	(0.0)	0.011	(84.1)	0.891
EBIT/EV[1]	−0.46	(78.8)	1.133	(0.0)	0.355	(1.6)	−0.591	(0.0)	0.054	(38.2)	0.934
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2]	−0.32	(58.8)	0.937	(0.0)	−0.696	(0.0)	−0.141	(0.0)	−0.015	(42.0)	0.947
TOPSIS2 (D/P, EBITDA/EV)[2]	−0.51	(75.0)	1.155	(0.0)	0.333	(1.2)	−0.799	(0.0)	0.085	(13.3)	0.892
AHP4 (CF3/P, D/P, MOM, S/EV)[2]	−4.04	(0.3)	1.126	(0.0)	0.234	(5.6)	−0.662	(0.0)	−0.042	(41.2)	0.902
MS4 (CFO/P, D/P, MOM, S/EV)[2]	−2.29	(8.1)	1.115	(0.0)	0.223	(0.0)	−0.354	(0.0)	0.026	(43.2)	0.916
D/P[2]	−1.94	(7.0)	1.084	(0.0)	−0.198	(0.2)	−0.104	(2.1)	0.062	(4.8)	0.924

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within either each of the four combination methods or among the 13 single selection criteria. Significances (in percentages) for the regression coefficients are in parentheses.

AHP, whereas the median-scaled top combinations are the last in both cases.

According to the GRS test statistic, the joint test of zero decile alphas is rejected for all the top combinations, except for add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM). The most significant GRS test statistic is reported for TOPSIS (D/P, EBITDA/EV). By contrast, the null of non-monotonically decreasing decile alphas holds for all the top combinations (at the 5% significance level), other than for the same add.DEA criterion, for which the null of jointly zero decile alphas is not rejected. This finding tangibly highlights the robustness differences between the GRS test and the MR test for decile alphas.

In order to find out the reasons for the differences between the results on the relative discriminatory power of MCDM methods being compared, we next examine the factor exposures of the top- and bottom-decile portfolios of the eight top combination criteria. Table 5 Panel A shows the corresponding regression slopes (MKT (i.e., the market factor), SMB, HML, and WML) for the top-decile portfolios, whereas Panel B does the same for the bottom-decile portfolios (for comparability purposes, the corresponding statistics are also shown for the best single selection criteria). Among the extreme top-combination deciles, the MKT slopes range from 0.842 of the top-decile AHP4 (CF3/P, D/P, MOM, S/EV) portfolio to 1.155 of the bottom-decile TOPSIS2 (D/P, EBITDA/EV) portfolio. However, the lowest overall MKT slope (0.775) is documented for the top-decile D/P portfolio. Generally, the MKT slopes are clearly higher for the bottom-decile portfolios than they are for their top-decile counterparts. The only exceptions are the two add.DEA criteria, for which the reverse holds. As expected, the market beta is positive and extremely significant for all decile portfolios examined.

With respect to SMB exposures, there is much more diversity: among the top-decile portfolios, four out of the eight top-combination SMB slopes are significantly positive, indicating that these four portfolios are tilted towards the small-cap stocks, whereas the top-decile MS4 (CFO/P, D/P, MOM, S/EV) portfolio is the only one having a significant large-cap tilt. Both add.DEA top-decile (bottom-decile) portfolios have very significant positive (negative) exposure to the SMB factor. Thus, both these DEA-based criteria select small-cap stocks in the top-decile portfolios, and large-cap stocks in the bottom deciles, whereas the reverse holds for the MS4 (CFO/P, D/P, MOM, S/EV) criterion. Regarding the top-decile portfolios formed on the AHP and TOPSIS top combinations, those that are the best in terms of raw returns have a significant small-cap exposure, whereas those chosen on the basis of the Sharpe ratio have negative, albeit insignificant, SMB slopes.

The value factor (HML) is significant in 14 out of 16 extreme-decile portfolios formed on the top combination criteria. The slopes have the expected signs (i.e., positive for the top-decile portfolios and negative for the bottom-decile portfolios), except that the top-decile add.DEA3 (i: EV; o: BE, DIV, MOM) portfolio has a significantly negative HML slope, indicating a tilt towards glamour stocks rather than value stocks. The same add.DEA criterion is also exceptional in the sense that among the bottom-decile portfolios, it is the only one without a significantly negative HML exposure. In addition, the corresponding top-decile portfolio has a positive exposure on the momentum factor (which is significant at the 5.7% level), whereas all the other top-decile portfolios formed on the top combinations, except that based on the other add.DEA criterion, have a negative momentum exposure, which is significant in three out of eight cases. In order to find the reason why the top-decile add.DEA portfolios would have so different momentum exposures than the other top combinations, we examined the cross-sectional distributions of each selection criterion at each portfolio-formation point and found that they were remarkably narrower for the momentum criterion than for the other selection criteria. Thus, compared to the three other methods employed, it seems that the add.DEA would give more weight on the selection criteria for which the cross-sectional measures of dispersion, such as standard deviation, range and pseudo-standard deviation (defined as quartile range divided by 1.35), are relatively low. Overall, the momentum factor (WML) is less important for the sample portfolios than the other factors. This is not surprising given the facts that our combinations can include only one momentum indicator but many value indicators, and in addition, that the momentum factor employed in the 4-factor model is based on 12-month price momentum with one-month lag, and the monthly reformation of the winner 30% and loser 30% portfolios, whereas our

Table 6
Average decile returns for the top selection criteria and momentum criterion in bearish months.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
D/P[2]	−2.23	−2.83	−2.71	−2.96	−3.33	−3.34	−3.34	−3.80	−3.99	−4.62	2.38
MS4 (CFO/P, D/P, MOM, S/EV)[2]	−2.46	−2.99	−3.15	−3.46	−4.27	−4.71	−4.86	−4.95	−5.19	−5.54	3.08
AHP4 (CF3/P, D/P, MOM, S/EV)[2]	−2.47	−3.01	−3.26	−3.76	−4.04	−4.27	−4.60	−4.77	−5.27	−6.32	3.85
TOPSIS2 (D/P, EBITDA/EV)[2]	−2.60	−2.99	−3.11	−3.36	−3.90	−3.97	−4.63	−4.95	−5.88	−6.27	3.67
EBIT/EV[1]	−3.24	−3.38	−3.18	−3.50	−3.66	−3.78	−4.30	−4.95	−5.62	−5.95	2.71
MS4 (B/P, EBITDA/EV, MOM, S/EV)[1]	−3.32	−3.54	−3.41	−3.45	−3.77	−4.01	−4.29	−4.66	−5.07	−6.08	2.76
TOPSIS2 (EBITDA/EV, S/P)[1]	−3.35	−3.40	−3.29	−3.38	−3.52	−3.84	−4.28	−4.60	−5.81	−6.11	2.77
AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1]	−3.65	−3.79	−3.86	−3.68	−3.89	−3.90	−3.93	−4.15	−4.73	−6.00	2.34
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2]	−3.83	−4.27	−4.37	−4.38	−4.31	−4.26	−4.33	−4.13	−4.08	−3.61	−0.22
MOM	−3.98	−3.88	−3.56	−3.27	−3.27	−3.54	−4.08	−4.61	−5.27	−6.18	2.20
add.DEA3 (i: EV; o: BE, DIV, MOM)[1]	−4.06	−4.49	−4.36	−4.54	−4.26	−4.07	−4.35	−4.03	−3.92	−3.36	−0.70

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within either each of the four combination methods or among the 13 single selection criteria.

Table 7
Average decile returns for the top selection criteria and momentum criterion in bullish months.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
add.DEA3 (i: EV; o: BE, DIV, MOM)[1]	5.03	4.86	4.65	4.58	4.54	4.46	4.52	4.23	4.14	3.59	1.44
MOM	4.87	4.78	4.32	4.13	4.07	4.12	4.16	4.51	4.76	4.94	−0.07
add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2]	4.80	4.68	4.77	4.63	4.59	4.55	4.42	4.26	4.16	3.71	1.08
AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1]	4.80	4.48	4.46	4.40	4.46	4.37	4.41	4.33	4.34	4.59	0.21
TOPSIS2 (EBITDA/EV, S/P)[1]	4.65	4.36	4.14	4.09	4.18	4.13	4.37	4.51	5.01	5.26	−0.61
MS4 (B/P, EBITDA/EV, MOM, S/EV)[1]	4.61	4.42	4.37	4.22	4.19	4.37	4.39	4.54	4.74	4.80	−0.19
EBIT/EV[1]	4.42	4.17	4.11	4.13	4.09	4.21	4.41	4.70	5.08	5.36	−0.94
TOPSIS2 (D/P, EBITDA/EV)[2]	3.97	3.98	4.09	4.19	4.34	4.30	4.48	4.74	5.19	5.43	−1.46
AHP4 (CF3/P, D/P, MOM, S/EV)[2]	3.82	3.86	4.19	4.21	4.46	4.57	4.73	4.85	5.01	4.90	−1.07
MS4 (CFO/P, D/P, MOM, S/EV)[2]	3.64	3.84	3.98	4.09	4.43	4.90	4.78	4.98	5.05	4.95	−1.31
D/P[2]	3.40	3.67	3.62	3.87	3.94	3.83	4.07	4.07	4.28	4.54	−1.14

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within either each of the four combination methods or among the 13 single selection criteria.

momentum indicator is based on equally lagged 6-month price momentum and a 12-month holding period for decile portfolios (in addition, the portfolio breakpoints for the WML factor are set so that each of the momentum decile portfolios has an equal number of NYSE firms (i.e., French, 2015, uses NYSE breakpoints), which reduces the winner-loser decile return spread compared to the practice in which the deciles are formed purely on the basis of the decile breakpoints of past return rankings (see Daniel, 2015, for further details). However, as the WML factor is significant for 16 out of the 80 top-combination decile portfolios (being significantly positive (negative) in 9 (7) cases), it is important that it is included in the regression model.

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 divide the full sample period into bullish and bearish months based on the sign of the market return, in line with Fuller and Goldstein (2011).[16] Based on that criterion, the full 42-year sample period includes 308 bullish and 196 bearish months. Table 6 shows the geometric average monthly returns during the bearish months for the decile portfolios formed on the top combination criteria augmented with the best single selection criteria (i.e., EBIT/EV in terms of raw returns and D/P in terms of the Sharpe ratio) and the momentum criterion. Among the eight top combinations included, 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, although the plain top-decile D/P portfolio provides the best hedge in such conditions (with an average return of −2.23%). The decomposition reveals that the return accumulation of the two add.DEA-based top-decile portfolios differs a lot from that of their other MCDM counterparts: during the bearish months, the former portfolios generate the highest losses, whereas during the bullish months (Table 7), they earn the highest returns. By contrast, the lowest average losses among the eight top combination criteria during the bearish months are generated by those MS-, AHP-, and TOPSIS-based combinations (in descending order) that are the best in terms of the Sharpe ratio. During the bullish months, the same combinations earn the lowest returns within the same peer-group in reverse order compared to the corresponding bearish-month return ranking. In fact, the rank orders based on bearish and bullish months (shown in Tables 6 and 7) are totally reversed.

Tables 6 and 7 also include the average bearish and bullish month returns for the pure momentum portfolios to demonstrate that the two add.DEA-based top-decile portfolios behave similarly to the top-decile momentum portfolio as they are the greatest losers during the bearish months while also the biggest winners during the bullish months. Hence, given that the top-decile MS4 (CFO/P, D/P, MOM, S/EV) portfolio that is the best top combination criteria during the bearish months and the worst during the bullish months also includes momentum as one sub-criterion, the add.DEA combination criteria tend to give greater weight to momentum than the MS-based criterion in portfolio selection. This finding is also supported by the WML slopes reported in Table 5: for the two add.DEA-based top-decile portfolios, they are positive, while they are negative for all their counterpart portfolios formed on the basis of the other MCDM criteria. In addition, the average top-bottom decile return spread is positive during the bearish months for all other top combinations, except for the add.DEA combinations. By

[16] 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, Biscarri, and de Gracia (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 (available from the authors upon request) are qualitatively similar to those based on average returns of bullish and bearish months.

Table 8
Sub-period raw return rankings for the top-decile portfolios formed on the 13 single selection criteria.

Sub-period	B/P	CF/P	CF2/P	CF3/P	CFO/P	D/P	E/P	EBIT/EV	EBITDA/EV	FCF/EV	MOM	S/EV	S/P
Panel A: rankings over the 7-year sub-periods													
5/1971–4/1978	4	5	7	6	8	3	10	2	1	12	9	11	13
5/1978–4/1985	13	7	9	4	8	11	10	3	5	12	6	2	1
5/1985–4/1992	12	7	8	10	6	13	9	3	5	11	1	2	4
5/1992–4/1999	8	9	11	4	6	13	12	5	3	2	1	7	10
5/1999–4/2006	10	7	4	6	9	13	12	5	3	11	8	1	2
5/2006–4/2013	9	8	10	7	5	2	6	4	3	12	13	11	1
Panel B: rankings over the 3-year sub-periods													
Sub-period	B/P	CF/P	CF2/P	CF3/P	CFO/P	D/P	E/P	EBIT/EV	EBITDA/EV	FCF/EV	MOM	S/EV	S/P
5/1971–4/1974	4	3	7	8	6	1	10	9	5	13	2	12	11
5/1974–4/1977	3	10	11	6	5	9	4	1	2	8	13	7	12
5/1977–4/1980	12	3	5	1	8	13	10	9	4	11	2	6	7
5/1980–4/1983	10	6	7	5	11	12	13	4	8	9	3	1	2
5/1983–4/1986	13	9	10	11	3	6	2	1	5	12	8	7	4
5/1986–4/1989	2	1	3	5	10	13	12	9	7	11	8	4	6
5/1989–4/1992	11	9	13	12	6	2	8	5	7	3	1	4	10
5/1992–4/1995	5	4	8	7	6	13	12	2	1	3	11	9	10
5/1995–4/1998	11	9	10	5	6	13	8	7	4	2	1	12	3
5/1998–4/2001	4	11	9	8	12	5	13	7	6	3	2	1	10
5/2001–4/2004	11	4	3	6	7	13	9	5	8	12	10	1	2
5/2004–4/2007	11	3	4	2	5	13	8	7	6	12	9	10	1
5/2007–4/2010	9	6	7	5	2	8	3	1	4	11	13	12	10
5/2010–4/2013	4	6	7	8	10	2	11	9	5	12	13	3	1

contrast, it is positive during the bullish months for the add.DEA combinations, but negative for all other top combinations, except for AHP4 (CF3/P, FCF/EV, MOM, S/EV) that generates positive top-bottom decile return spreads in both bearish and bullish months, thereby lending further support to its feasibility as the basis of a long-short investment strategy (the same combination criteria also generates the highest, as well as the most significant, top-bottom alpha spread).

As the next robustness check, we examine whether the top-decile rankings of selection criteria have any consistency over time by dividing the full-length sample period into sub-periods of equal lengths. We divide the 42-year sample period into six 7-year sub-periods and 14 3-year sub-periods, because for such sub-period divisions, there is only one common portfolio-reformation point, and the lengths of sub-periods are reasonably different to each other for the evaluation of medium- and short-term performance persistence. The multi-period consistency between the sub-period rankings is evaluated on the basis of Kendall's coefficient of concordance (also known as Kendall's W). We first restrict this test to include only the rankings based on single selection criteria, owing to the arduousness of generating rankings for a huge number of combinations with time-consuming and computationally burdensome combination methods. Table 8 shows the corresponding raw return rankings for each of the 7-year (Panel A) and 3-year (Panel B) sub-periods. The coefficient of concordance for the 7-year periods is significant at the 1.48% level, but insignificant for the 3-year periods (significance level is 15.05%) indicating that the consistency of the sub-period raw return rankings is higher for the longer sub-periods, at least in this particular case and for the sub-period lengths employed.[17]

We also repeated the same tests by adding the eight top-decile portfolios formed on the top combinations in the calculation of Kendall's W, bearing in mind that this type of test design is prone to look-ahead bias (The rankings are shown in Appendix Table A1). The results showed that the raw return ranking consistency remains approximately the same for the 7-year sub-periods (significant at the 2.22% level), but improves for the 3-year sub-periods (significant at the 1.23% level) in a consequence of adding the top combinations besides the 13 single selection criteria into the analysis. Similar, but more intensive increase in significance levels is documented for the top-decile rankings formed on the Sharpe ratio, based on which the coefficient of concordance is extremely significant (even at the 0.05% level) for both sub-period divisions.

Although the above-reported results show some evidence of ranking consistency over the sub-periods, Table 8 and Appendix Table A1 show that the rankings vary a lot between consecutive sub-periods. Therefore, we also calculated the performance for the top-decile portfolio by updating the selection criteria to match the one that generated the highest returns during the preceding sub-period. This enabled us to compare the performance between the top-decile portfolios formed on stable and varying selection criteria over the 39-year (35-year) holding period in case of 3-year (7-year) sub-periods. For the sample of top-decile portfolios formed on the 13 single selection criteria, the raw return of the top-decile portfolio formed on the best past selection criterion (determined in terms of raw returns) is rather close to the average performance of the corresponding portfolios formed on the single selection criteria: For 3-year (7-year) sub-periods, the average raw return for the first-mentioned portfolio over the period from May 1974 (May 1978) to April 2013, is 18.72% p.a. (18.16% p.a.), whereas the average raw return for the 13 top-decile portfolios formed on the single selection criteria is 18.50% p.a. (17.38% p.a.), respectively.[18]

[17] We also repeated the similar test for rank concordance by determining the best past selection criterion based on one-year historical returns and found that Kendall's W is insignificant in that case. In addition, we made the similar type of comparison by determining the best past selection criterion based on the top-decile Sharpe ratios calculated over the previous 3-year (7-year) period. In that case, the rank concordance for the 7-year sub-periods is significant at the 5% level, whereas for the 3-year sub-periods, it is weakly significant (at the 10% level). The detailed results are available from the authors upon request.

[18] When the best past selection criterion was determined on the basis of the top-decile Sharpe ratios over the previous 3-year (7-year) period, the raw return of the corresponding top-decile portfolio was slightly lower compared to that generated by determining the best past selection criterion in terms of raw returns (18.48% p.a. for 3-year sub-periods and 17.96% p.a. for 7-year sub-periods). By contrast, in terms of Sharpe ratios, the reverse relation held, indicating that determining the selec-

Table 9
Average decile returns for the combination methods: the large-cap sample.

Panel A: averages for 2-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	14.65	13.88	12.79	12.23	11.86	11.34	10.88	9.56	9.73	8.21	6.44
MS	14.52	13.82	13.07	12.39	11.72	11.33	10.43	9.66	9.73	8.44	6.07
AHP	14.51	13.76	13.31	12.27	11.76	11.30	10.74	9.70	9.26	8.50	6.01
add.DEA (i: ME)	13.70	12.81	13.62	11.22	10.92	10.93	11.51	11.05	9.93	9.42	4.28
add.DEA (i: EV)	13.60	13.50	12.01	10.49	11.27	11.21	12.08	11.35	10.30	9.30	4.31
Panel B: averages for 3-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	15.00	13.75	13.02	12.29	11.70	11.26	10.95	9.73	9.24	8.17	6.83
MS	14.81	13.58	13.12	12.41	11.73	11.38	10.75	9.55	9.44	8.34	6.47
AHP	14.63	13.91	13.41	12.59	11.69	11.26	10.77	9.56	9.00	8.31	6.32
add.DEA (i: ME)	14.27	12.89	13.11	11.97	10.70	10.99	11.02	10.91	10.02	9.23	5.04
add.DEA (i: EV)	14.19	13.33	12.41	10.78	11.15	10.85	11.65	11.27	10.28	9.21	4.98
Panel C: averages for 4-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	15.16	13.80	13.20	12.33	11.51	11.09	11.04	9.84	9.25	7.90	7.27
MS	15.01	13.34	13.49	12.34	11.56	10.97	11.16	9.74	9.51	8.00	7.00
AHP	14.82	13.69	13.47	12.55	11.63	11.46	10.95	9.45	9.10	7.99	6.83
add.DEA (i: ME)	14.71	13.07	12.58	12.17	11.04	11.00	10.93	10.69	9.83	9.10	5.61
add.DEA (i: EV)	14.68	13.31	12.32	11.17	11.04	10.51	11.57	11.03	10.39	9.10	5.58
add.DEA (i: EV, ME)	14.22	13.15	12.87	10.62	10.80	10.94	11.69	11.26	10.17	9.36	4.86

Note: The decile averages are reported separately for the add.DEA combinations that use market equity (ME), enterprise value (EV), or both (ME & EV) as input(s).

However, for 3-year (7-year) sub-periods, six (five) out of 13 top-decile portfolios formed on the single selection criteria generated higher returns than the top-decile portfolio formed on the best past selection criterion.

4.4. The results for the large-cap sample

Besides our main sample, we examine the large-cap sample that consists of the firms whose market equity value is higher than the NYSE breakpoint of the second highest size quintile. This robustness check is motivated by the earlier results, according to which both momentum and value anomalies are partially explained by size effect (e.g., Hong, Lim, & Stein, 2000; Lesmond, Schill, & Zhou, 2004 report that the profitability of momentum portfolios declines with firm size. By contrast, according to Israel & Moskowitz, 2013, momentum profits exhibit no reliable relation with size, but value premium is largely concentrated only in small stocks and is insignificant among 40% of the largest-cap stocks. However, they drew the last-mentioned conclusion on the basis of B/P as the only value indicator. Motivated by the results of Israel & Moskowitz, 2013, Pätäri, Karell, Luukka, & Yeomans, 2017 test whether the value premiums determined on the basis of valuation ratios other than B/P behave in the same way in the similarly designed sample of large-cap stocks. Interestingly, the authors find that the value premium is remarkably higher based on many other valuation ratios, particularly based on S/EV, for which it is more than double that based on B/P. Actually, based on the return difference between extreme decile portfolios during the 1971–2013 sample period, the B/P-based value premium is the lowest among all 12 valuation ratios examined. However, the results of these studies are not totally comparable due to differences not only in their research designs, but also in the sample periods employed. Nevertheless, in light of these results, it is interesting to check whether the added-value from using combinations remains or declines among large-cap stocks included in the two biggest size quintiles (i.e., stocks above 40% NYSE market-cap breakpoint)).

For the large-cap sample, the results are somewhat different to those for the main sample. First, the average decile returns and the corresponding Sharpe ratios for each combination method are more even for the comparable deciles indicating that the differences in the discriminatory power between the combination methods is attenuated when the sample consists of only large-cap stocks (see Tables 9 and 10). In this case, the average performance statistics for the MS combinations behave similarly to those for other combination methods, i.e., the average returns of the MS-based top-decile portfolios, as well as the corresponding top-bottom decile return spreads, increase monotonically when moving from 2-combinations to 3-combinations, and likewise when moving from 3-combinations to 4-combinations, whereas the reverse holds for the main sample. The similar pattern is also documented in terms of the Sharpe ratios. Nevertheless, the TOPSIS method generates both the highest average top-decile returns and the highest average Sharpe ratios, as it also does in case of the main sample, but with narrower margin in this case. For the large-cap sample, it also produces the highest average top-bottom decile spreads in terms of both raw returns and the Sharpe ratios, whereas for the main sample, the corresponding highest spreads are generated by the AHP.

Table 11 shows the annualized geometric average returns (Panel A) and the Sharpe ratios (Panel B) for the top combination criteria that, within each of the four combination methods generate the highest top-decile rankings based on either of these two perfor-

tion criterion based on the past raw returns seems to result in somewhat higher return but lower Sharpe ratio than in the case where it was chosen on the basis of the Sharpe ratio. However, the differences are small in terms of both raw and risk-adjusted returns, and the added-value of changing the selection criteria periodically is negative in many cases. We also tested whether more frequent updating of the best past selection criterion would have paid off to the investor by determining the best past selection criterion based on one-year historical returns. The results showed that the relative performance of such a selection criterion was worse than for 3-year and 7-year updating periods, as ten out of 13 top-decile portfolios formed on the single selection criteria generated higher returns than the top-decile portfolio formed on the best past selection criterion. In terms of risk-adjusted returns, its relative performance was even worse as its Sharpe ratio was lower than for any of the 13 last-mentioned top-decile portfolios.

Table 10
Average Sharpe ratios for the combination methods: the large-cap sample.

Panel A: averages for 2-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	0.525	0.508	0.442	0.408	0.380	0.334	0.298	0.219	0.192	0.121	0.405
MS	0.520	0.504	0.457	0.414	0.372	0.331	0.275	0.225	0.195	0.128	0.392
AHP	0.508	0.487	0.459	0.406	0.374	0.333	0.291	0.229	0.181	0.130	0.378
add.DEA (i: ME)	0.474	0.402	0.423	0.323	0.305	0.305	0.329	0.318	0.260	0.253	0.221
add.DEA (i: EV)	0.435	0.412	0.359	0.287	0.321	0.316	0.366	0.342	0.289	0.258	0.177
Panel B: averages for 3-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	0.549	0.504	0.459	0.415	0.372	0.331	0.298	0.225	0.175	0.118	0.431
MS	0.541	0.497	0.465	0.421	0.372	0.332	0.284	0.216	0.185	0.123	0.418
AHP	0.528	0.496	0.463	0.419	0.367	0.330	0.290	0.224	0.174	0.124	0.404
add.DEA (i: ME)	0.506	0.410	0.404	0.356	0.299	0.303	0.305	0.310	0.260	0.243	0.263
add.DEA (i: EV)	0.470	0.404	0.376	0.304	0.317	0.297	0.344	0.335	0.284	0.254	0.216
Panel C: averages for 4-combinations											
Combination method	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
TOPSIS	0.558	0.509	0.470	0.420	0.364	0.324	0.300	0.229	0.175	0.109	0.449
MS	0.556	0.488	0.487	0.418	0.368	0.315	0.296	0.220	0.186	0.112	0.444
AHP	0.548	0.492	0.472	0.421	0.362	0.334	0.294	0.216	0.179	0.115	0.433
add.DEA (i: ME)	0.529	0.420	0.383	0.364	0.314	0.300	0.301	0.298	0.251	0.238	0.291
add.DEA (i: EV)	0.501	0.400	0.369	0.322	0.314	0.285	0.339	0.324	0.287	0.250	0.250
add.DEA (i: EV, ME)	0.473	0.404	0.392	0.298	0.297	0.302	0.344	0.341	0.282	0.258	0.215

Note: The decile averages are reported separately for the add.DEA combinations that use market equity (ME), enterprise value (EV), or both (ME & EV) as input(s).

Table 11
Decile portfolio performance for the top combination and the best single selection criteria: the large-cap sample.

Panel A: raw returns											
Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
add.DEA4 (i: EV; o: CFO, DIV, EBITDA, MOM)*	17.70	12.44	12.14	11.89	10.98	10.69	10.48	9.97	10.55	8.28	9.43
	(0.1)	(97.3)	(64.8)	(64.6)	(15.7)	(16.2)	(18.9)	(17.6)	(52.8)	(7.9)	(0.0)
MS3 (CF3/P, MOM, S/EV)	17.18	12.27	13.81	13.11	12.57	12.04	9.99	10.34	8.29	5.52	11.66
	(0.4)	(70.4)	(11.6)	(15.2)	(37.3)	(88.6)	(7.4)	(11.3)	*(0.4)*	(0.0)	(0.0)
AHP3 (B/P, CF3/P, S/EV)	17.18	11.38	13.28	14.81	10.90	9.65	11.32	9.61	9.55	7.44	9.73
	(0.2)	(66.8)	(18.2)	(0.4)	(44.3)	*(1.8)*	(44.9)	*(2.5)*	(6.2)	*(0.5)*	(0.0)
TOPSIS3 (CF3/P, MOM, S/EV)	16.74	13.51	12.43	13.11	11.91	11.68	12.97	8.92	7.98	5.85	10.89
	(1.1)	(8.2)	(44.5)	(26.7)	(95.7)	(89.4)	(54.3)	*(0.9)*	*(0.3)*	*(0.0)*	(0.0)
S/EV***	16.15	14.47	13.30	11.84	10.19	11.35	10.82	9.81	9.15	8.03	8.12
	(18.8)	(44.9)	(32.1)	(83.4)	(12.9)	(57.2)	(47.9)	(18.1)	(9.0)	*(1.6)*	(2.1)
Panel B: Sharpe ratios											
Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10
AHP4 (CF2/P, D/P, EBIT/EV, S/EV)	0.660	0.485	0.468	0.300	0.364	0.396	0.270	0.243	0.113	0.093	0.567
	(0.3)	(10.5)	(26.6)	(37.6)	(92.4)	(51.3)	(17.1)	(6.7)	*(0.0)*	*(0.1)*	(0.0)
MS4 (CF2/P, D/P, EBITDA/EV, S/EV)	0.635	0.471	0.475	0.400	0.368	0.272	0.308	0.185	0.169	0.079	0.556
	(0.8)	(22.3)	(24.4)	(67.9)	(96.8)	(14.9)	(55.5)	*(1.7)*	*(0.8)*	*(0.1)*	(0.0)
TOPSIS3 (CF2/P, D/P, S/EV)**	0.635	0.479	0.405	0.444	0.356	0.360	0.290	0.167	0.197	0.063	0.572
	(0.6)	(14.5)	(62.4)	(34.8)	(93.6)	(88.8)	(28.6)	*(0.2)*	*(2.4)*	*(0.0)*	(0.0)
add.DEA4 (i: EV; o: CFO, DIV, EBITDA, MOM)*	0.608	0.370	0.334	0.342	0.291	0.286	0.283	0.274	0.314	0.193	0.415
	(0.1)	(97.3)	(64.8)	(64.6)	(15.7)	(16.2)	(18.9)	(17.6)	(52.8)	(7.9)	(0.0)
D/P***	0.563	0.544	0.515	0.430	0.346	0.294	0.331	0.237	0.258	0.185	0.377
	(4.5)	(5.4)	(14.8)	(70.5)	(52.3)	(11.3)	(23.6)	*(0.7)*	(7.5)	*(2.0)*	(0.8)

Note: Panels also indicate the significance levels (in percentages in parentheses) for the Sharpe ratio difference between each decile portfolio and the market portfolio. The last column shows the corresponding top-bottom decile (D1–D10) differences, respectively (All these significance levels are based on the Ledoit–Wolf test and italicized in case of significant underperformance). The significances for the MR test for decile returns are shown by asterisks. * (**/***) indicates that decile returns are monotonically decreasing at the 1% (5%/10%) significance level. MS refers to median-scaling, i: (o:) refers to input (output) variable(s) included in the add.DEA method, DIV to dividends, and MOM to momentum.

mance metrics (for the large-cap sample, the number of top combinations is seven instead of eight because the best add.DEA combination is the same in terms of both raw returns and the Sharpe ratio). In contrast to the main sample results, the highest decile return among the large-cap top combinations (17.70%) is documented for the top-decile portfolio formed on the add.DEA (i: EV; o: CFO, DIV, EBITDA, MOM). This return is surprisingly high as the excess return of this particular portfolio over the sample average (that is 11.52%) is even slightly higher for this sample (6.18%) than for the corresponding excess return for the main sample (6.01%) despite the fact that based on financial literature, pricing efficiency has been documented to be higher among large-caps than among

Table 12
4-factor decile alphas for the top combination and the best single selection criteria: the large-cap sample.

Portfolio-formation criterion	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D1–D10	GRS
add.DEA4 (i: EV; o: CFO, DIV, EBITDA, MOM)[1,2]	4.60	−0.94	−0.71	−0.91	−0.72	−1.24	−0.61	−1.04	0.15	−0.36	4.96	4.22
	(0.0)	(36.6)	(50.3)	(32.4)	(42.5)	(18.6)	(47.8)	(17.8)	(85.2)	(71.3)	(0.2)	(0.0)
S/EV[1] **	4.15	2.90	1.32	0.28	−1.38	−0.63	−1.40	−1.26	−2.41	−2.86	7.01	1.65
	(0.4)	(2.8)	(18.2)	(77.3)	(13.1)	(41.4)	(12.6)	(14.2)	*(1.9)*	*(4.3)*	(0.1)	(9.1)
AHP3 (B/P, CF3/P, S/EV)[1]	3.52	−1.19	0.22	1.70	−0.36	−2.00	0.01	−1.05	−0.44	−2.01	5.53	4.35
	(0.1)	(15.3)	(79.3)	(7.6)	(71.1)	*(1.8)*	(99.4)	(22.6)	(68.3)	(9.4)	(0.1)	(0.0)
MS3 (CF3/P, MOM, S/EV)[1]	3.05	−1.15	0.65	−0.27	−0.31	0.16	−1.01	0.33	−1.14	−0.86	3.91	1.74
	(0.5)	(20.0)	(45.0)	(74.7)	(74.1)	(85.2)	(24.1)	(73.5)	(25.6)	(60.3)	(4.8)	(6.9)
AHP4 (CF2/P, D/P, EBIT/EV, S/EV)[2]	2.78	0.22	0.43	−1.59	−0.13	1.35	−0.05	0.25	−1.52	−1.90	4.68	1.52
	(1.3)	(81.4)	(64.0)	(8.9)	(88.9)	(16.6)	(95.8)	(79.7)	(17.5)	(16.6)	(0.8)	(13.0)
TOPSIS3 (CF3/P, MOM, S/EV)[1]	2.55	0.08	−0.97	0.13	−0.94	−0.99	1.84	−1.48	0.11	−0.50	3.05	1.55
	(1.7)	(92.3)	(27.9)	(88.2)	(26.1)	(24.8)	(5.1)	(10.0)	(91.4)	(77.1)	(13.2)	(12.0)
MS4 (CF2/P, D/P, EBITDA/EV, S/EV)[2]	2.46	0.05	0.20	−0.74	−0.38	−1.17	1.34	0.30	−0.59	0.09	2.37	2.76
	(4.0)	(95.5)	(82.7)	(44.4)	(66.6)	(19.7)	(20.7)	(79.3)	(58.1)	(95.8)	(23.8)	(0.3)
TOPSIS3 (CF2/P, D/P, S/EV)[2] **	2.32	0.38	−0.33	0.16	−0.61	0.30	0.07	−1.03	−0.67	−0.42	2.75	0.61
	(5.4)	(71.7)	(70.7)	(85.3)	(47.9)	(73.3)	(93.8)	(30.3)	(53.7)	(77.0)	(14.5)	(80.5)
D/P[2] ***	1.22	0.77	1.06	−0.32	−0.91	−1.87	−0.82	−2.64	−1.29	−1.36	2.58	0.99
	(34.6)	(44.2)	(37.0)	(77.4)	(38.0)	(9.5)	(42.1)	*(1.1)*	(29.4)	(26.8)	(14.8)	(44.9)

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within either each of the four combination methods or among the 13 single selection criteria. Significances (in percentages) for the alphas and for the top-bottom decile alpha spreads (D1–D10) are in parentheses and italicized in case of significant underperformance. The last column shows the GRS statistics (significance levels in percentages in parentheses). The significances for the MR test for decile alphas are shown by asterisks. ** (***) indicates that decile alphas are monotonically decreasing at the 5% (10%) significance level.

more heterogeneous-sized firms (see e.g., Lo & MacKinlay, 1990; Griffin & Lemmon, 2002; Hou & Moskowitz, 2005). Interestingly, the corresponding top-decile return of the best TOPSIS combination is the lowest (16.74%) among the four combination methods, whereas it was the highest for the main sample. Nevertheless, all the four highest-return top combinations generate higher raw returns than the best decile portfolios formed on single selection criteria (among the decile portfolios based on the single selection criteria, the highest raw return (16.15%) within the large-cap sample is documented for the top-decile S/EV portfolio). In addition, the incremental return from using the best combination criterion instead of the best single selection criteria is somewhat higher for the large-cap sample than for the main sample (1.55% vs. 1.25% p.a.), indicating that in terms of raw returns, the benefits from combinations are, at their best, even greater in the former sample.

Based on the MR test, decile returns decrease monotonically and significantly for two out of the seven top combination criteria. The corresponding test statistics is the most significant (at the 1% level) for the add.DEA 4-combination criterion, followed by TOPSIS3 (CF2/P, D/P, S/EV), for which it is significant at the 5% level. Thus, in terms of return differences between the deciles, these two combination criteria have the best discriminatory power throughout the deciles.

In terms of the Sharpe ratio, the top-decile rank order of the combination methods is different to the corresponding raw return ranking (Table 11, Panel B). The highest Sharpe ratio (0.660), which is quite close to the corresponding maximum for the main sample (i.e. 0.688), is documented for the top-decile AHP (CF2/P, D/P, EBIT/EV, S/EV). By contrast, the same add.DEA criterion that is the best in terms of top-decile raw returns, generates the lowest top-decile Sharpe ratio (0.608) among the four comparable combination methods (the Sharpe ratio for the equal-weighted benchmark portfolio that consists of the stocks above 40% NYSE market-cap breakpoint is 0.340). Also for the large-cap sample, the risk-adjusted results look different on the basis of the 4-factor alphas (Table 12): the highest and the most significant 4-factor alpha (4.60%) is documented for the same add.DEA-based top-decile portfolio that also generates the highest raw return among all decile portfolios. The comparison of the maximum alphas over the samples reinforces that abnormal returns that would have been achievable based on the best combination criteria are approximately at the same level in the large-sap sample than they are in the main sample. Thus, the risk-adjusted top-decile performance of the best combination criteria is surprisingly similar in both samples, indicating that slightly lower returns of the best large-cap portfolios are, at their best, almost fully compensated by their slightly lower risk. The factor exposures (not reported here but available from the authors upon request) are also very similar in both samples. For example, the above-discussed top-decile add.DEA large-cap portfolio has a significant small-cap tilt, as well as a significant momentum exposure, whereas its bottom-decile counterpart is significantly tilted towards large-caps with significantly negative momentum exposure.

The GRS test statistic shows that the decile alphas are, on average, closer to zero in the large-cap sample than in the main sample, as the null hypothesis of jointly zero alphas is rejected (at the 5% significance level) only for three out the seven top combinations, whereas among the main-sample top combinations, the corresponding rejection rate was seven out of eight. For the large-cap sample, the null of the MR test, according to which the decile alphas are not monotonically decreasing, holds for all the top combination criteria, except for TOPSIS3 (CF2/P, D/P, S/EV). By contrast, the GRS test statistic for this particular combination is insignificant, thereby highlighting the differences between the GRS test and the alpha-based MR test.

Table 13 shows the occurrence frequencies of each single selection criterion in the top combination criteria for both the main and large-cap samples. For the latter, the top combinations are more coherent. For example, the S/EV criterion is included in all the top-combination criteria other than that based on the best add.DEA method. Among those top combinations selected on the basis of the Sharpe ratio, D/P and CF2/P are also included in all other than the DEA-based top combination, whereas among the top combinations based on the highest raw returns, the same holds for the CF3/P. Among the latter top combinations, TOPSIS and MS combinations include exactly the same selection criteria. Based on these results, it seems that when the sample of stocks is more homogeneous, the divergence between the AHP-, MS-, and TOPSIS-based combination methods on the best variable combinations diminishes. By contrast, add.DEA deviates from other methods also in this sense. However, it should be noted that the feasible variable

Table 13
Occurrence frequencies of the single selection criteria in the top combinations.

Portfolio-formation criterion	Main sample	Large-cap sample	Total
B/P	2 (1/1)	1 (1/0)	3 (2/1)
CF1/P	0	0	0
CF2/P	0	3 (0/3)	3 (0/3)
CF3/P	2 (1/1)	3 (3/0)	5 (4/1)
CFO/P	1 (0/1)	0	1 (0/1)
D/P	4 (0/4)	3 (0/3)	7 (0/7)
E/P	0	0	0
EBIT/EV	0	1 (0/1)	1 (0/1)
EBITDA/EV	3 (2/1)	2 (1/2)	6 (3/3)
FCF/EV	1 (1/0)	0	1 (1/0)
MOM	6 (3/3)	3 (3/1)	10 (6/4)
S/EV	4 (2/2)	6 (3/3)	10 (5/5)
S/P	1 (1/0)	0	1 (1/0)

Note: In parentheses are given the proportion of occurrences in the top combinations that, within each combination method, have generated the highest raw return to occurrences in those, which have generated the highest Sharpe ratios, respectively. Because for the large-cap sample, the best add.DEA combination is the same in terms of both raw returns and Sharpe ratios, its sub-criteria have been double-counted in occurrence frequencies anywhere else in the table than in the total numbers reported separately for the large-cap sample.

combinations are somewhat different in case of the add.DEA, for the reason discussed in Section 4.1.

5. Discussion

Overall results show that the MCDM methods can add value to equity portfolio selection. Although the outperformance of the best top-decile portfolios formed on the basis of top combination criteria over their counterpart portfolios formed on the basis of single selection criteria is not statistically significant, it is economically significant. For example, the cumulative return difference between the highest-yield decile portfolio formed on the basis of the best single selection criterion (i.e., EBIT/EV) and the corresponding portfolio formed on the basis of the top combination criterion (i.e., TOPSIS2 (EBITDA/EV and S/P)) is approximately 56.34% during the 42-year holding period, implying that without transaction costs, the one-dollar initial investment would have grown to a \$973 terminal value in the former case, whereas in the latter case, the terminal value would have ended up at \$1.521, approximately. For the large-cap sample, the corresponding cumulative return difference is even higher (75.14%) as the corresponding terminal values are approximately \$536 (for the S/EV top-decile portfolio) and \$939 (for the top-decile portfolio formed on the add.DEA (i: EV; o: CFO, DIV, EBITDA, MOM).

Like most academic peer-group studies, we have not included transaction costs in our analysis because their level is both investor- and trade-specific (see, e.g., Keim & Madhavan, 1997; Lewellen, 2010). Although their exclusion creates 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, Israel, & Moskowitz, 2015; Novy-Marx & Velikov, 2016). 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 the rebalancing trades rather than the sale or purchase of total stockholdings are needed. It is also unlikely that the added value of combining single selection criteria into combination criteria would decrease if the transaction costs were included in the analysis because the turnover in the portfolios formed on the combination criteria is hardly higher than it is in the portfolios formed on a single selection criterion.[19] Therefore, the inclusion of transaction costs would have only a marginal impact on comparisons of the relative performance of decile portfolios, although it would slightly decrease the statistical significance of the outperformance of the best portfolios while increasing the significance of the underperformance of the worst. In addition, the inclusion of transaction costs can slightly influence the relative performance of the best portfolios, as these portfolios have divergent market-cap exposures, and the price impact from implementing trades is stronger among smaller-cap stocks (see, e.g., Chiyachantana, Jain, Jiang, & Wood, 2004).

The constituent stocks included in each decile portfolio were equally-weighted at annual portfolio-formation points, but weight changes stemming from return differences during the one-year holding periods were taken into account in the calculation of the monthly time-series returns of the 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 footnote 8 for the motivation behind the chosen methodology). However, the results for the value-weighted returns could have been different from ours, although an additional robustness check, where the sample was restricted to include only the largest U.S. stocks above the second highest NYSE market-cap quintile breakpoint, showed that the benefits of combining single selection criteria remained qualitatively the same.

Besides value and price momentum indicators, other selection criteria have also been shown to enhance equity portfolio performance. For example, the benefits of including profitability measures in value portfolio selection are 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, since the focus of our analysis is on the value and momentum dimensions, these kinds of extensions are beyond the scope of this study, although they provide an interesting subject for further research. With respect to the combination methods employed, we concentrate on relatively simple variations of MCDM methods that are applicable to creating combination rankings. The results of this study do not exclude the possibility that the benefits of combining could have been greater had we used either more sophisticated or simpler combination methods.[20] Therefore, one possible extension for further research could be to examine whether more sophisticated portfolio-formation approaches based on optimization techniques could add value to equity portfolio selection (see, e.g., Aouni, Colapinto, & La Torre, 2014; Ben Abdelaziz, Aouni, & El Fayedh, 2007; Dia, 2009; Ehrgott, Ide, & Schöbel, 2014; Fliege & Werner, 2014; Kim, Kim, Kim, & Fabozzi, 2014; Kiris & Ustun, 2012; Kolm, Tütüncü, & Fabozzi, 2014; Mansini, Ogryczak, & Speranza, 2014; Ogryczak, 2000; Xidonas, Mavrotas, & Psarras, 2009; Xidonas, Mavrotas, Zopounidis, & Psarras, 2011; Xidonas, Mavrotas, Hassapis, Zopounidis, 2017).

The results reported in this study are based on a simple backtesting approach and they may be specific to the sample period. Based on experimental sub-period tests, there is some concor-

[19] Actually, the recent results of Novy-Marx and Velikov (2016) would indicate that the reverse could rather hold, as they report lower transaction costs for the combined value-momentum strategy than for the pure momentum strategy. However, the portfolios in their comparison are updated monthly, unlike in our study.

[20] As examples of other MCDM methods used for equity portfolio selection, Albadvi, Chaharsooghi, and Esfahanipour (2007) use the PROMETHEE, whereas Huck (2009) and Xidonas et al. (2009) use the ELECTRE Tri (Both of these methods are classified as outranking techniques). 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.

dance in top-decile performance rankings within the full sample period. Of course, this does not necessarily imply that the same would also hold for the future, as the out-of-sample results are similarly specific to the sample period. Nevertheless, the comprehensive out-of-sample tests would provide an interesting topic for further research, although they are beyond the scope of this paper. For example, a similar type of multicriteria framework as employed in this study could be extended towards building predictive models.

6. Conclusions

We test a larger universe of combination strategies than examined in any of the previous studies by employing many MCDM methods that had never been used before for forming value-momentum equity portfolios. The results for both samples show that an equity investor can benefit from combining single selection criteria into one composite measure. The inclusion of a momentum indicator besides value measures also enhances the performance of combination portfolios. With respect to the discriminatory power of the portfolio formation, there are differences between the combination methods. At least for this particular sample data and in terms of raw and total risk-adjusted returns, the AHP- and TOPSIS-based combinations are able to separate the future outperforming stocks from their underperforming counterparts somewhat better than the median-scaled combinations or the add.DEA-based combinations. However, for the main sample, the highest 4-factor alpha, as well as the most significant return- and alpha-based MR test statistics among the eight top combinations are documented for top-decile add.DEA portfolios. In addition, for the large-cap sample, the best-performing portfolio in terms of both raw returns and 4-factor alphas is also a top-decile add.DEA portfolio, indicating that an equity investor should not overlook the add.DEA as a potential combination method.

The variable combinations that are the best in terms of either raw or total risk-adjusted returns are rarely the same, although there is some consistency in what variables are most commonly included in the best combination strategies. The clearest evidence is documented for the additional value of the inclusion of D/P, stemming from its ability to reduce portfolio risk and enhance the total risk-adjusted portfolio performance. Furthermore, the benefits from the inclusion of one or more EV multiple(s) are also evident in most cases, as is the inclusion of a momentum indicator. Of EV multiples, S/EV is most often included in the top combinations.

The division of the full 42-year sample period into bullish and bearish months reveals more differences between the top combination criteria that are based on different MCDM methods: the top-decile add.DEA portfolios behave similarly to the top-decile momentum portfolio earning the highest returns during the bullish months, while losing the most during the bearish months, whereas the reverse holds for the MS- and AHP-based top-decile portfolios. The tendency of add.DEA to give greater weight to the momentum indicator in stock selection than is given by the other MCDM methods being compared makes add.DEA distinct from the others. Although the MCDM methods employed in this study lead to different factor exposures, the results show that these methods can also add value to equity portfolio formation in the most developed and efficient stock markets in the world. Importantly, the added value also extends to the sample of large-cap stocks implying that the benefits from combination are not restricted to small-scale investors. As a practical implication of this study, investors following certain investing styles could take these different style exposures into account when choosing the MCDM criteria that best suit their portfolio-selection purposes.

Acknowledgements

We thank the editor (Emanuele Borgonovo), two anonymous referees, as well as Jukka Perttunen and other seminar participants at the Graduate School of Finance Summer Workshop 2016 at the University of Vaasa for their valuable comments and suggestions. We are also grateful to Teemu Toivonen for excellent research assistance.

Appendix

Table A1
Sub-period raw return rankings for the top-decile portfolios formed on the 13 single selection criteria and the top combinations.

Panel A: rankings over the 7-year sub-periods

Sub-period	B/P	CF/P	CF2/P	CF3/P	CFO/P	D/P	E/P	EBIT/EV	EBITDA/EV	FCF/EV	MOM	S/EV	S/P	add.DEA3[1]	add.DEA4[2]	AHP4[1]	AHP4[2]	MS4[1]	MS4[2]	TOPSIS2[1]	TOPSIS2[2]
5/1971–4/1978	9	11	14	13	15	7	18	6	5	20	17	19	21	1	12	16	2	3	4	10	8
5/1978–4/1985	21	15	17	12	16	19	18	11	13	20	14	8	7	9	6	1	5	2	4	3	10
5/1985–4/1992	19	14	15	17	13	21	16	9	12	18	2	7	11	3	20	1	6	4	8	5	10
5/1992–4/1999	14	15	17	10	12	21	19	11	5	4	1	13	16	3	2	9	20	6	8	7	18
5/1999–4/2006	16	13	10	12	15	20	19	11	9	17	14	2	4	8	3	7	18	6	21	1	5
5/2006–4/2013	12	11	13	10	8	3	9	6	4	19	20	16	2	21	18	15	1	17	14	7	5

Panel B: rankings over the 3-year sub-periods

Sub-period	B/P	CF/P	CF2/P	CF3/P	CFO/P	D/P	E/P	EBIT/EV	EBITDA/EV	FCF/EV	MOM	S/EV	S/P	add.DEA3[1]	add.DEA4[2]	AHP4[1]	AHP4[2]	MS4[1]	MS4[2]	TOPSIS2[1]	TOPSIS2[2]
5/1971–4/1974	10	9	13	15	12	1	18	16	11	21	6	20	19	4	7	14	2	8	3	17	5
5/1974–4/1977	5	18	19	14	13	17	11	1	3	16	21	15	20	7	12	10	9	2	8	4	6
5/1977–4/1980	18	6	8	4	13	21	16	15	7	17	5	11	12	1	3	10	19	9	14	2	20
5/1980–4/1983	18	13	14	12	19	20	21	11	15	16	10	3	7	4	6	1	2	5	9	8	17
5/1983–4/1986	20	15	16	17	8	12	3	2	11	18	14	13	10	21	19	5	9	1	4	7	6
5/1986–4/1989	5	4	6	8	16	21	20	13	10	19	12	7	9	18	17	3	11	2	15	1	14
5/1989–4/1992	19	17	21	20	13	6	16	11	15	8	4	10	18	1	12	7	2	9	5	14	3
5/1992–4/1995	7	6	10	9	8	21	18	3	1	5	16	13	14	12	11	17	20	2	19	4	15
5/1995–4/1998	16	14	15	8	10	21	12	11	7	4	1	19	6	13	9	3	20	5	17	2	18
5/1998–4/2001	7	19	17	16	20	8	21	10	9	5	4	3	18	1	2	15	12	14	13	11	6
5/2001–4/2004	18	8	7	11	12	20	15	10	13	19	16	3	6	17	9	1	14	2	21	4	5
5/2004–4/2007	15	4	5	3	9	18	12	11	10	17	13	14	1	20	19	8	16	7	21	2	6
5/2007–4/2010	14	7	10	6	2	12	3	1	4	17	19	18	15	21	16	11	9	20	13	5	8
5/2010–4/2013	8	12	13	14	18	3	19	16	11	20	21	4	1	17	10	15	2	6	5	9	7

Note: Superscript 1 (2) indicates the portfolio-formation criterion that generates the highest top-decile raw return (Sharpe ratio) within each of the four combination methods in the full sample period. Constituents for the top combinations: add.DEA3 (i: EV; o: BE, DIV, MOM)[1], add.DEA4 (i: ME; o: BE, DIV, EBIT, MOM)[2], AHP4 (CF3/P, FCF/EV, MOM, S/EV)[1], AHP4 (CF3/P, D/P, MOM, S/EV)[2], MS4 (B/P, EBITDA/EV, MOM, S/EV)[1], MS4 (CFO/P, D/P, MOM, S/EV)[2], TOPSIS2 (EBITDA/EV, S/P)[1], and TOPSIS2 (D/P, EBITDA/EV)[2].

References

Albadvi, A., Chaharsooghi, S. K., & Esfahanipour, A. (2007). Decision making in stock trading: An application of PROMETHEE. *European Journal of Operational Research, 177*, 673–683.

Aouni, B., Colapinto, C., & La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. *European Journal of Operational Research, 234*, 536–545.

Asness, C. S. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal, 53*, 29–36.

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

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

Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2007). Momentum and credit rating. *Journal of Finance, 63*, 2503–2520.

Banker, R. D., Chen, J. Y. S., & Klumpes, P. (2016). A trade-level DEA model to evaluate relative performance of investment fund managers. *European Journal of Operational Research, 255*, 903–910.

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics, 116*, 111–120.

Bartov, E., & Kim, M. (2004). Risk, mispricing, and value investing. *Review of Quantitative Finance and Accounting, 23*, 353–376.

Beaver, W., McNichols, M., & Price, R. (2007). Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics, 43*, 341–368.

Behzadian, M., Otaghsara, S. K., Yazdani, M., & Ignatius, J. (2012). A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications, 39*, 13051–13069.

Ben Abdelaziz, F., Aouni, B., & El Fayedh, R. (2007). Multi-objective stochastic programming for portfolio selection. *European Journal of Operational Research, 177*, 1811–1823.

Billio, M., Calès, L., & Guégan, D. (2011). Portfolio symmetry and momentum. *European Journal of Operational Research, 214*, 759–767.

Bird, R., & Casavecchia, L. (2007a). Sentiment and financial health indicators for value and growth stocks: The European experience. *European Journal of Finance, 13*, 769–793.

Bird, R., & Casavecchia, L. (2007b). Value enhancement using momentum indicators: The European experience. *International Journal of Managerial Finance, 3*, 229–262.

Bouyssou, D. J. (1999). Using DEA as a tool for MCDM: Some remarks. *Journal of the Operational Research Society, 50*, 974–978.

Cakici, N., Fabozzi, F. J., & Tan, S. (2013). Size, value, and momentum in emerging market stock returns. *Emerging Markets Review, 16*, 46–65.

Cakici, N., & Tan, S. (2014). Size, value, and momentum in developed country equity returns: Macroeconomic and liquidity exposures. *Journal of International Money and Finance, 44*, 179–209.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance, 52*, 57–82.

Chan, L. K. C., Hameed, A., & Tong, W. (2000). Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis, 35*, 153–172.

Chan, L. K. C., Jegadeesh, N., & Lakonishok, J. (1996). Momentum strategies. *Journal of Finance, 51*, 1681–1713.

Chan, L. K. C., & Lakonishok, J. (2004). Value and growth investing: Review and update. *Financial Analysts Journal, 60*, 71–86.

Chang, C.-H., Lin, J.-J., Lin, J.-H., & Chiang, M.-C. (2010). Domestic open-end equity mutual fund performance evaluation using extended TOPSIS method with different distance approaches. *Expert Systems with Applications, 37*, 4642–4649.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L. M., & Stutz, J. (1985). Foundations of data envelopment analysis and Pareto–Koopmans empirical production functions. *Journal of Econometrics, 30*, 91–107.

Chen, Y., Li, K. W., & Liu, S.-F. (2011). An OWA-TOPSIS method for multiple criteria decision analysis. *Expert Systems with Applications, 38*, 5205–5211.

Chiyachantana, C. N., Jain, P. K., Jiang, C., & Wood, R. A. (2004). International evidence on institutional trading behavior and price impact. *Journal of Finance, 59*, 869–898.

Chui, A. C. W., Titman, S., & Wei, K. C. J. (2010). Individualism and momentum around the world. *Journal of Finance, 65*, 361–392.

Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research, 192*, 1–17.

Cooper, M., Gutierrez, R. C., Jr., & Hameed, A. (2004). Market states and momentum. *Journal of Finance, 59*, 1345–1365.

Daniel, K. D. (2015). Notes on the Daniel–Moskowitz momentum portfolios. Available at: http://www.kentdaniel.net/data/momentum/mom_data.pdf.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance, 55*, 389–406.

Desai, H., Rajgopal, S., & Venkatachalam, M. (2004). Value-glamour and accruals mispricing: One anomaly or two? *Accounting Review, 79*, 355–385.

Dhatt, M. S., Kim, Y. J., & Mukherji, S. (2004). Can composite value measures enhance portfolio performance? *Journal of Investing, 13*, 42–48.

Dia, M. (2009). A portfolio selection methodology based on data envelopment analysis. *Information Systems and Operational Research, 47*, 71–79.

Doyle, J., & Green, R. (1993). Data envelopment analysis and multiple criteria decision making. *Omega, 21*, 713–715.

Du, J., Chen, C.-M., Chen, Y., Cook, W. D., & Zhu, J. (2012). Additive super-efficiency in integer-valued data envelopment analysis. *European Journal of Operational Research, 218*, 186–192.

Dymova, L., Sevastjanov, P., & Tikhonenko, A. (2013). A direct interval extension of TOPSIS method. *Expert Systems with Applications, 40*, 4841–4847.

Edwards, S., Biscarri, J. G., & de Gracia, F. P. (2003). Stock market cycles, financial liberalization and volatility. *Journal of International Money and Finance, 22*, 925–955.

Ehrgott, M., Ide, J., & Schöbel, A. (2014). Minmax robustness for multi-objective optimization problems. *European Journal of Operational Research, 239*, 17–31.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance, 47*, 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*, 3–56.

Fama, E. F., & French, K. R. (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics, 60*, 3–43.

Fama, E. F., & French, K. R. (2006a). Profitability, investment and average return. *Journal of Financial Economics, 82*, 491–518.

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

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

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics, 105*, 457–472.

Figelman, I. (2007). Stock return momentum and reversal. *Journal of Portfolio Management, 34*, 51–67.

Fisher, G., Shah, R., & Titman, S. (2016). Combining value and momentum. *Journal of Investment Management, 14*, 33–48.

Fliege, J., & Werner, R. (2014). Robust multiobjective optimization & applications in portfolio optimization. *European Journal of Operational Research, 234*, 422–433.

Fong, W. M., & Ong, Z. (2016). The profitable dividend yield strategy. *Journal of Investment Management, 14*, 51–61.

Frazzini, A., Israel, R., & Moskowitz, T. (2015). *Trading costs of asset pricing anomalies*. IL: University of Chicago. Unpublished working paper.

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

Fuller, K. P., & Goldstein, M. A. (2011). Do dividends matter more in declining markets? *Journal of Corporate Finance, 17*, 457–473.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica, 57*, 1121–1152.

Gray, W. R., & Vogel, J. (2012). Analyzing valuation measures: A performance horse-race over the past 40 years. *Journal of Portfolio Management, 39*, 112–121.

Gutierrez, R. C., & Kelley, E. K. (2008). The long-lasting momentum in weekly returns. *Journal of Finance, 63*, 415–447.

Griffin, J. M., & Lemmon, M. L. (2002). Book-to-market equity, distress risk, and stock returns. *Journal of Finance, 57*, 2317–2336.

Ho, C.-T. B., & Oh, K. B. (2010). Selecting Internet company stocks using a combined DEA and AHP approach. *International Journal of Systems Science, 41*, 325–336.

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research, 202*, 16–24.

Hong, H., Lim, T., & Stein, J. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance, 55*, 265–295.

Hou, K., Karolyi, G. A., & Kho, B. C. (2011). What factors drive global stock returns? *Review of Financial Studies, 24*, 2527–2574.

Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies, 18*, 981–1020.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies, 28*, 650–705.

Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research, 196*, 819–825.

Hwang, C.-L., & Yoon, K. (1981). *Multiple attributes decision making methods and applications*. Berlin, Heidelberg: Springer.

Hwang, S.-N., Lin, C-T., & Chuang, W-C. (2007). Formulating a novel stock selection model using DEA and grey situation decision model. *Journal of Statistics and Management Systems, 10*, 103–127.

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications, 38*, 14336–14345.

Israel, R., & Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics, 108*, 275–301.

Israelsen, C. L. (2005). A refinement to the Sharpe ratio and information ratio. *Journal of Asset Management, 5*, 423–427.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*, 255–259.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance, 48*, 65–91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance, 56*, 699–720.

Jones, C. S., & Pomorski, L. (2017). Investing in disappearing anomalies. *Review of Finance, 21*, 237–267.

Keim, D. B., & Madhavan, A. (1997). Execution costs and investment style: An inter-exchange analysis of institutional equity trades. *Journal of Financial Economics, 46*, 265–292.

Kim, W. C., Kim, M. J., Kim, J. H., & Fabozzi, F. J. (2014). Robust portfolios that do not tilt factor exposure. *European Journal of Operational Research, 234*, 411–421.

Kiris, S., & Ustun, O. (2012). An integrated approach for stock evaluation and portfolio optimization. *Optimization, 61*, 423–441.

Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research, 234*, 356–371.

Korajczyk, R., & Sadka, R. (2004). Are momentum profits robust to trading costs? *Journal of Finance, 59*, 1039–1082.

Kosowski, R., Naik, N. Y., & Teo, M. (2007). Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics, 84*, 229–264.

Kou, G., & Lin, C. (2014). A cosine maximization method for the priority vector derivation in AHP. *European Journal of Operational Research, 235*, 225–232.

Kuo, T. (2017). A modified TOPSIS with a different ranking index. *European Journal of Operational Research, 260*, 152–160.

Lakonishok, J., Shleifer, A., & Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance, 49*, 1541–1578.

Lam, K., Liu, T., & Wong, W.-K. (2010). A pseudo-Bayesian model in financial decision making with implications to market volatility, under- and overreaction. *European Journal of Operational Research, 203*, 166–175.

Ledoit, O., & 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.

Leivo, T. H., & Pätäri, E. J. (2011). Enhancement of value portfolio performance using momentum and the long-short strategy: The Finnish evidence. *Journal of Asset Management, 11*, 401–416.

Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics, 71*, 349–380.

Lewellen, J. (2010). Accounting anomalies and fundamental analysis: An alternative view. *Journal of Accounting and Economics, 50*, 455–466.

Lin, C.-C. (2007). A revised framework for deriving preference values from pairwise comparison matrices. *European Journal of Operational Research, 176*, 1145–1150.

Lo, A. W., & MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies, 3*, 175–205.

Loughran, T., & 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.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2014). Twenty years of linear programming based portfolio optimization. *European Journal of Operational Research, 234*, 518–535.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance, 71*, 5–32.

Newey, W. K., & West, K. D. (1987). A simple semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*, 703–708.

Ogryczak, W. (2000). Multiple criteria linear programming model for portfolio selection. *Annals of Operations Research, 97*, 143–162.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics, 103*, 429–453.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108*, 1–28.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies, 29*, 104–147.

Palkar, D. D., & Wilcox, S. E. (2009). Adjusted earnings yields and real rates of return. *Financial Analysts Journal, 65*, 66–79.

Patton, A. J., & Timmermann, A. (2010). Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts. *Journal of Financial Economics, 98*, 605–625.

Penman, S., & Reggiani, F. (2013). Returns to buying earnings and book value: Accounting for growth and risk. *Review of Accounting Studies, 18*, 1021–1049.

Pätäri, E. J., Karell, V., & 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.

Pätäri, E. J., Karell, V., Luukka, P., & Yeomans, J. S. (2017). The dirty dozen of valuation ratios: Is one better than the other? *Journal of Investment Management* (Forthcoming).

Pätäri, E. J., & 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., & 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., & Honkapuro, J. V. S. (2012). Enhancement of equity portfolio performance using data envelopment analysis. *European Journal of Operational Research, 220*, 786–797.

Rouwenhorst, K. G. (1998). International momentum strategies. *Journal of Finance, 53*, 267–284.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology, 15*, 234–281.

Saaty, T. L., Rogers, P. C., & Pell, R. (1980). Portfolio selection through hierarchies. *Journal of Portfolio Management, 6*, 16–21.

Saaty, T. L., & Shang, J. (2011). An innovative orders-of-magnitude approach to AHP-based multicriteria decision making: Prioritizing divergent intangible humane acts. *European Journal of Operational Research, 214*, 703–715.

Sahoo, B. K., Khoveyni, M., Eslami, R., & Chaudhury, P. (2016). Returns to scale and most productive scale size in DEA with negative data. *European Journal of Operational Research, 255*, 545–558.

Scowcroft, A., & Sefton, J. A. (2005). Understanding momentum. *Financial Analysts Journal, 61*, 64–82.

Sevastjanov, P., & Dymova, L. (2009). Stock screening with use of multiple criteria decision making and optimization. *Omega, 37*, 659–671.

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.

Stewart, T. J. (1994). Data envelopment analysis and multiple criteria decision making: A response. *Omega, 22*, 205–206.

Stewart, T. J. (1996). Relationships between data envelopment analysis and multicriteria decision analysis. *Journal of the Operational Research Society, 47*, 654–665.

Tiryaki, F., & Ahlatcioglu, M. (2005). Fuzzy stock selection using a new fuzzy ranking and weighting algorithm. *Applied Mathematics and Computation, 170*, 144–157.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research, 130*, 498–509.

Treynor, J. L., & Black, F. (1973). How to use security analysis to improve portfolio selection. *Journal of Business, 46*, 66–86.

Walkshäusl, C., & Lobe, S. (2015). The enterprise multiple investment strategy: International evidence. *Journal of Financial and Quantitative Analysis, 50*, 781–800.

Xidonas, P., Mavrotas, G., Hassapis, C., & Zopounidis, C. (2017). Robust multiobjective portfolio optimization: A minimax regret approach. *European Journal of Operational Research, 262*, 299–305.

Xidonas, P., Mavrotas, G., & Psarras, J. (2009). A multicriteria methodology for equity selection using financial analysis. *Computers and Operations Research, 36*, 3187–3203.

Xidonas, P., Mavrotas, G., Zopounidis, C., & Psarras, J. (2011). IPSSIS: An integrated multicriteria decision support system for equity portfolio construction and selection. *European Journal of Operational Research, 210*, 398–409.

Zhu, B., Xu, Z., Zhang, R., & Hong, M. (2016). Hesitant analytic hierarchy process. *European Journal of Operational Research, 250*, 602–614.

Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., & Andriosopoulos, K. (2015). Multiple criteria decision aiding for finance: An updated bibliographic survey. *European Journal of Operational Research, 247*, 339–348.