

Yeh, I-Cheng; Liu, Yi-Cheng

## Article

# Discovering optimal weights in weighted-scoring stock-picking models: A mixture design approach

Financial Innovation

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Yeh, I-Cheng; Liu, Yi-Cheng (2020) : Discovering optimal weights in weighted-scoring stock-picking models: A mixture design approach, Financial Innovation, ISSN 2199-4730, Springer, Heidelberg, Vol. 6, Iss. 1, pp. 1-28,  
<https://doi.org/10.1186/s40854-020-00209-x>

This Version is available at:

<https://hdl.handle.net/10419/237228>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

METHODOLOGY

Open Access



# Discovering optimal weights in weighted-scoring stock-picking models: a mixture design approach

I-Cheng Yeh<sup>1\*</sup> and Yi-Cheng Liu<sup>2</sup>

\*Correspondence:  
140910@mail.tku.edu.tw

<sup>1</sup> Department of Civil  
Engineering, Tamkang  
University, 151 Yingzhuang  
Rd., Tamsui Dist., New Taipei  
City 25137, Taiwan  
Full list of author information  
is available at the end of the  
article

## Abstract

Certain literature that constructs a multifactor stock selection model adopted a weighted-scoring approach despite its three shortcomings. First, it cannot effectively identify the connection between the weights of stock-picking concepts and portfolio performances. Second, it cannot provide stock-picking concepts' optimal combination of weights. Third, it cannot meet various investor preferences. Thus, this study employs a mixture experimental design to determine the weights of stock-picking concepts, collect portfolio performance data, and construct performance prediction models based on the weights of stock-picking concepts. Furthermore, these performance prediction models and optimization techniques are employed to discover stock-picking concepts' optimal combination of weights that meet investor preferences. The samples consist of stocks listed on the Taiwan stock market. The modeling and testing periods were 1997–2008 and 2009–2015, respectively. Empirical evidence showed (1) that our methodology is robust in predicting performance accurately, (2) that it can identify significant interactions between stock-picking concepts' weights, and (3) that which their optimal combination should be. This combination of weights can form stock portfolios with the best performances that can meet investor preferences. Thus, our methodology can fill the three drawbacks of the classical weighted-scoring approach.

**Keywords:** Portfolio optimization, Stock-picking, Weighted-scoring, Mixture experimental design, Multivariable polynomial regression analysis

## Highlights

- Finding the connection between weights of stock-picking concepts and performances.
- Discovering the optimal combination of weights of stock-picking concepts.
- Meeting various investors' preferences.

## Introduction

The efficient-market hypothesis asserts that financial markets reflect all relevant information. Consequently, stocks always trade at their fair value on stock exchanges, making it impossible for investors to either purchase undervalued stocks or sell overvalued stocks. Certain empirical studies show that stock markets do not reach semi-strong market efficiency (Hong and Stein 1999; Hong et al. 2000; Piotroski 2000; Richardson et al. 2010; Fama and French 2012; Asness et al. 2013; Yeh and Hsu 2014; Kong et al. 2019; Daniel et al. 2020; Wen et al. 2019). Banz's (1981) size effect suggests that the return on investment (ROI) in stocks of small corporations is higher than that in stocks of large corporations. Rosenberg et al.'s (1985) value effect indicates that value stocks have higher ROI than growth stocks. Bondt and Thaler's (1985) overreaction and reversal effects illustrate that winner stocks have a lower ROI than loser stocks do in the long term. The results of the momentum effect observed by Jegadeesh and Titman (1993) show that rising asset prices increase further, whereas falling prices keep decreasing. Thus, stocks with strong past performance will continue outperforming stocks with poor past performance during the next period.

Several recent studies have shown that a combination of effects could be used to construct a stock selection model with a high rate of return (RoR) (Piotroski 2000; Hart et al. 2003, 2005; Mohanram 2005; Qian et al. 2007; Roko and Gilli 2008; Yeh and Hsu 2011; Shen et al. 2015; Yu et al. 2014; Yeh et al. 2015; Rasekhschaffe and Jones 2019; Dai and Zhou 2019; Wu et al. 2020; Gu et al. 2020). Many studies also adopted a weighted-scoring approach to construct a multifactor stock selection model (Piotroski 2000; Kang and Ding 2006; Duran-Vazquez, et al. 2014; Kim and Lee 2014; Tikkanen and Äijö 2018; Jeong and Kim 2019; Mehta, et al. 2019). For example, Hart et al. (2003) studied the profitability of various stock selection strategies in 32 emerging markets from 1985 to 1999. Value, momentum, and earnings revision strategies were the most successful as they generated significant excess returns when compared with size, liquidity, and mean reversion strategies. A strategy can be improved efficiently by combining various stock-picking factors. Finally, large institutional investors can implement these strategies successfully regardless of liquidity constraints and significant transaction costs.

Mohanram (2005) proposed combining traditional fundamentals, such as earnings and cash flow, with company growth indicators, such as earnings stability, growth stability, R&D intensity, capital expenditures, and advertising, to establish an index, that is, a G-score. A long/short equity based on G-score generated significant excess returns although most of the returns were generated through shorting. The results were robust to in size, analyst coverage, and liquidity issues and persisted after controlling for momentum, price-to-book value ratios (P/B ratio), and accruals. Firms with a high G-score demonstrated strong market reaction to future earnings announcements and unannounced analyst forecast. In addition, a risk-based approach cannot explain the results because returns were positive in most years and low-risk companies earned high returns. Finally, fundamental analysis worked best when traditional and growth-based analyses were paired with stocks with high and low P/B ratio, respectively.

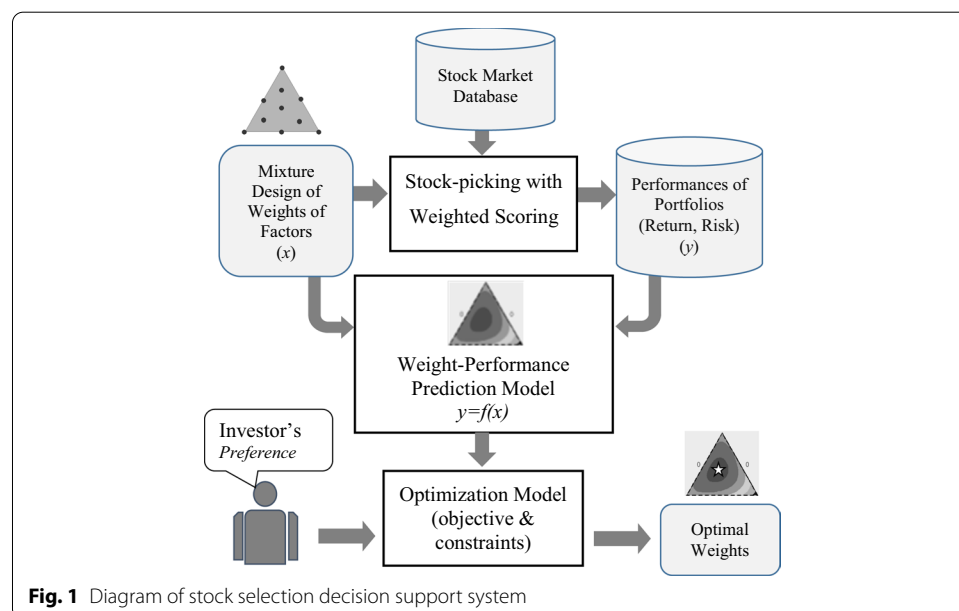
Noma (2010) combined traditional fundamentals, such as return on assets, operating cash flow, and operating margins, as an F-score index. F-score was applied, and it demonstrated that the mean return can increase by 7.8% through a hedging strategy

that buys high-F-score firms and that shorts firms with a low F-score. Additionally, an investment strategy that buys firms with a high P/B ratio and F-score and shorts those with a low P/B ratio and F-score earns a 17.6% annual return. The empirical result also reveals that the F-score can predict future earnings.

However, this method either sets up weights subjectively or uses a simple average, leading to three drawbacks. First, it cannot effectively identify the connection between the weights of stock-picking concepts and portfolio performances. Second, it cannot provide an optimal combination of stock-picking concepts' weights. Third, the method cannot meet various investor preferences. For example, a conservative investor may only withstand low risks; hence, returns would not be the first priority. Therefore, stock selection factor weights should differ from those of an aggressive investor who considers returns the top priority.

We address these shortcomings by adopting the following methodology (Fig. 1):

- (1) We design stock-picking concepts' weight combinations with a mixture design (Myers and Montgomery 2008; Montgomery 2012). Accordingly, we generate a set of weighted combinations ( $x$ ) of stock-picking concepts to collect information on performances with different weight combinations of stock-picking concepts.
- (2) Based on the mixture design, we simulate stock-picking concepts' weight combinations to obtain investment performances ( $y$ ) through backtesting using stock market trading historical database. These results can be collected and matched as ( $x$ ,  $y$ ) to construct a data set.
- (3) Based on the data set, we construct a performance prediction model,  $y=f(x)$ , by employing a multivariable polynomial regression analysis. The prediction model can examine the relationship between the performances and weights of stock-picking concepts and identify the interactions between concepts.



- (4) Based on the prediction model, we discover stock-picking concepts' optimal combination of weights that can form a stock portfolio with the best performance to meet investor preferences by way of optimization techniques.
- (5) We verify stock-picking concepts' optimal combination of weights through back-testing using stock market trading historical database to determine whether they can meet investor preferences.

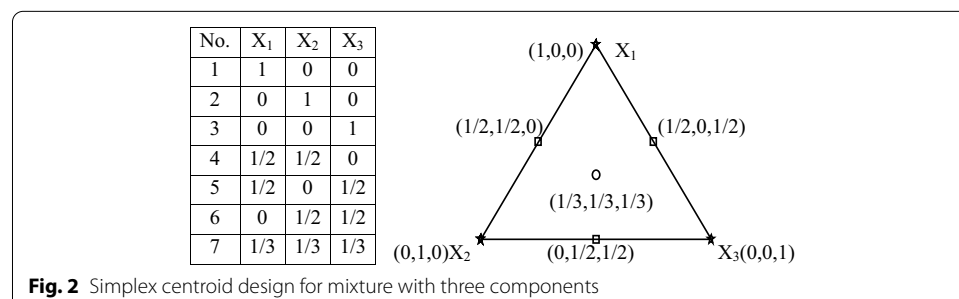
Therefore, this methodology can resolve the three drawbacks of the aforementioned extant literature. The remainder of this paper is organized as follows. “[Mixture experimental design](#)” section explains how we develop a mixture design. “[Experimental design and implementation](#)” section describes how we generate stock-picking concepts' weight combinations through this mixture experimental design and simulate them by way of backtesting. “[Model building and verification](#)” section constructs and analyzes the performance prediction model through a multivariable polynomial regression analysis. “[Weight optimization and validation](#)” section presents the determination of stock-picking concepts' optimal weight combinations through optimization and the validation of them through backtesting. Last, “[Conclusion](#)” section concludes the paper.

### Mixture experimental design

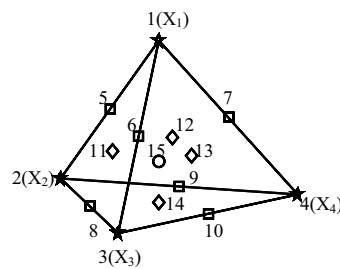
We systematically explore the relationship between different weight combinations of various factors and portfolio performances through mixture design, that is, a type of experimental designs, given that the sum of the weight combinations equals 1. The components of mixture experiments are their factors, the levels of which are dependent. Thus, each factor used in stock selection is assigned a weight, which is a level. We use the simplex-centroid design to conduct the mixture experimental design and thus produce various weight combinations (Montgomery 2012). The simplex-centroid design's  $q$ -type composition is expected to have  $2^q - 1$  experimental mixes. For example, Figs. 2 and 3 show the simplex-centroid design for a mixture with three and four components, respectively.

Then, we construct the model by employing a regression analysis using the experimental data obtained through backtesting. The polynomial functions of the simplex-centroid design are expressed as follows (Montgomery 2012):

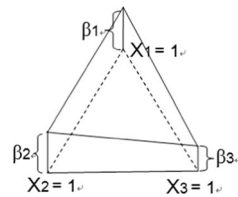
$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i<j}^q \sum \beta_{ij} x_i x_j + \sum_{i<j<k}^q \sum \sum \beta_{ijk} x_i x_j x_k + \cdots + \beta_{12\dots q} x_1 x_2 \dots x_q, \quad (1)$$



No.	$X_1$	$X_2$	$X_3$	$X_4$
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	1/2	1/2	0	0
6	1/2	0	1/2	0
7	1/2	0	0	1/2
8	0	1/2	1/2	0
9	0	1/2	0	1/2
10	0	0	1/2	1/2
11	1/3	1/3	1/3	0
12	1/3	1/3	0	1/3
13	1/3	0	1/3	1/3
14	0	1/3	1/3	1/3
15	1/4	1/4	1/4	1/4



**Fig. 3** Simplex centroid design for mixture with four components



**Fig. 4** The coefficient of the linear term is the regression estimates of the three apexes

where  $y$ ,  $x_i$ , and  $\beta$  represent the response variable of the mixture, the proportion of the  $i$ -th component of the mixture, and the regression coefficient of the regression model, respectively.

The effects of the higher-order terms can be ignored because they are usually small. In most real applications, only the first, second, and third terms may be significant. For instance, if  $q=3$  and three-order terms are included, then we have the following:

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 \quad (2)$$

According to Eq. (2) above, if the three-component mix is at the one-component mix (1, 0, 0), mix (0, 1, 0), or mix (0, 0, 1), then their expected responses are  $\beta_1, \beta_2$ , and  $\beta_3$ , respectively.

Figure 4 shows that the coefficient of the linear term is the regression estimates of the three apexes, and the average value of the coefficients of the three linear terms is the regression estimates of the central point. This finding indicates that the slope between the central point and apexes is positive if the coefficient of the linear term is larger than the average value of the coefficients of the three linear terms. Therefore, the correspondent regression estimate increases if the component is large. Conversely, the slope is negative if the coefficient is smaller than the average, the regression estimate decreases if the component is large.

If the three-component mix is at the two-component mix (1/2, 1/2, 0), (1/2, 0, 1/2), or (0, 1/2, 1/2), then according to Eq. (2), their respective expected responses are as follows:

$$E(y) = \frac{(\beta_1 + \beta_2)}{2} + \frac{\beta_{12}}{4} \quad (3)$$

$$E(y) = \frac{(\beta_1 + \beta_3)}{2} + \frac{\beta_{13}}{4} \quad (4)$$

$$E(y) = \frac{(\beta_2 + \beta_3)}{2} + \frac{\beta_{23}}{4} \quad (5)$$

Thus, the coefficient of quadratic term  $\beta_{ij}$  is four times the difference between the central-point regression estimate of side  $E(y)$  and average regression estimate values from the two apexes of side  $(\beta_i + \beta_j)/2$ . Therefore, the regression estimate of this side is a convex function if the coefficient of the quadratic term is greater than zero. Otherwise, the regression estimate of this side is a concave function if the coefficient of the quadratic term is less than zero (Fig. 5).

## Experimental design and implementation

### Factor Screening and stock-picking concepts

Stock selection factors can be divided into five categories as follows:

#### Value factors

Returns from cheap stocks are higher than those of expensive stocks. Commonly used ratios for these factors include the price-to-earnings (P/E) and P/B ratios.

#### Growth factors

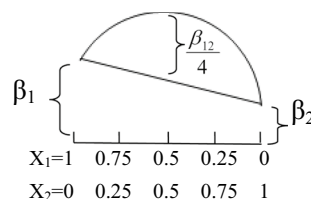
Stocks from profitable companies have higher returns than those from unprofitable companies do. Commonly used ratios include return on equity (ROE).

#### Momentum factors

Stocks with high recent returns have higher returns than those with low recent returns do. Quarterly and monthly RoRs are used to measure the momentum effect.

#### Size factors

Stocks from small firms have higher returns than those from large firms do. The total market capitalization (or market value) is commonly used to measure company size.



**Fig. 5** The coefficient of this quadratic term is four times the difference between the central-point regression estimate of the side and the average value of regression estimates from the two apexes of this side

### **Liquidity factors**

Stocks with low liquidity have higher returns than those with high liquidity do. Quarterly trading volume is commonly used to measure stock liquidity.

The performance indicators of portfolio investment can be divided into three categories, namely, returns, risks, and liquidity. Appropriate stock selection factors should be chosen to build a decision-making model that optimizes and satisfies these performance indicators as discussed below:

### **Returns**

The P/B ratio and ROE are the most representative indicators of value and growth stocks. In addition, the last quarterly and monthly RoRs may affect the RoR because of reversal or momentum effects in stock markets. This study uses the P/B ratio, ROE, and monthly RoR as stock selection factors.

Although the forward P/E ratio may depict the highest RoR for a portfolio, it is based on analysts' earnings forecasts. In fact, using appropriate combinations of weights, combining the P/B ratio and ROE (based on historical earnings), can achieve a RoR comparable with that achieved by the forward P/E ratio. Therefore, the forward P/E ratio was not used in this study because of the lack of evident advantages. Moreover, the P/B ratio and ROE represent stock value and growth, respectively, while the P/E ratio mixes value and growth. In terms of regulating the portfolio's various performance aspects, two independent factors are better than a single, mixed factor, which is another reason for not using the P/E ratio in this study.

### **Risks**

The  $\beta$  value of a stock is often continuous in nature; that is to say, stocks with large current  $\beta$  will usually have large future  $\beta$ , and vice versa. Therefore, previous  $\beta$  is chosen as a stock selection factor to control the systematic risks of the selected stocks.

### **Liquidity**

Market capitalization needs considerable time to grow or decline. Thus, stocks with large (or small) market value usually remain unchanged in the future. Although stocks with a small market value may generate a high return, certain investors may prefer investment targets that demonstrate significant liquidity and investment availability; hence, market value should not be extremely low. Therefore, stocks were ranked according to total market capitalization from large to small in order to ensure high market capitalization for selected stocks, whereby stocks with a large market value would have a corresponding high score.

This study employed the multifactor weighted method to select stocks that were sorted by default order according to stock selection factors. The scores for top- and bottom-end stocks were 100 and 0 points, respectively, whereas those for the remaining stocks were obtained through interpolation. Each factor's score was weighted to obtain the total weighted score. Furthermore, the stock with the highest total weighted score was considered the most profitable. Different weights form various stock-picking strategies and

have different performances. Hence, performance is a function of weights. Therefore, weights should be employed as design variables of the optimization model. Each factor also needs a default-sorting direction for the stocks as described below:

***Small P/B ratio concept***

The smaller the P/B ratio the higher the future returns. Therefore, stocks were sorted in ascending order according to the P/B ratio, with a smaller P/B ratio receiving a higher score.

***Large ROE concept***

The higher the ROE, the higher the future returns. Therefore, stocks were sorted in descending order according to ROE, with a higher ROE receiving a higher score.

***Large monthly return concept***

Stock market returns are usually characterized by short- and long-term reversals and middle-term momentum. Momentum effects typically occur during one or several months. Therefore, future stock returns may be high or low depending on the domination of either reversal or momentum effect, given a most recent high quarterly or monthly return. However, most investors psychologically prefer to buy stocks with high recent return. Therefore, stocks were sorted in descending order according to monthly returns, with a high monthly return receiving a corresponding high score.

***Large total market capitalization concept***

Total market value indicates company size. A company may be at the growing stage when it has low total market value. By contrast, a large total market capitalization implies that the company has established a leadership position in its industry. Although stocks of small firms may generate a high return, many investors may prefer investment targets that demonstrate significant liquidity and investment availability; hence, market value should not be extremely low. Therefore, stocks were sorted in descending order according to the total market capitalization to ensure that the selected stock had an appropriately high market capitalization, with a large total market capitalization receiving a corresponding high score.

***Small beta concept***

Beta ( $\beta$ ) measures stock return fluctuation relative to a benchmark (market), that is, systematic risks. A higher  $\beta$  implies that the stock return fluctuation is higher than that of the benchmark. If a stock's  $\beta$  is greater than 1, its return fluctuation is greater than the benchmark, and vice versa. A stock's beta value is often persistent; that is to say, stocks with large (small) current  $\beta$  would typically have large (small) future  $\beta$  in the near future. Although a large  $\beta$  may imply higher returns according to classic theory, many investors may prefer investment targets demonstrating significant low systematic risk; therefore,  $\beta$  cannot be too large. Hence, stocks were sorted in ascending order according to their  $\beta$  to reduce the selected stock's systematic risk. Stocks with small  $\beta$  receive high scores.

The weighted-scoring approach is employed to construct the weighted-scoring multi-factor stock selection model in two steps:

- (1) A single-factor scoring method is used to sort stocks. The top stock is assigned a score of 100, whereas the bottom stock is assigned a score of 0. An interpolation method is applied to the rest of stocks.
- (2) The multifactor scoring method is employed. We obtain each stock's overall score by assigning a certain weight to the score for each factor. Thus, the stock with the highest overall score is the best stock, and vice versa.

For example, previous literature results show that rate of return is high if the ROE is large and P/B ratio is small and vice versa. Therefore, the stock with the highest ROE or lowest P/B ratio is assigned a score of 100, whereas the stock with the lowest ROE or highest P/B ratio is assigned a score of 0. An interpolation method is applied to the rest of stocks. For example, a stock is assigned a score of 80 if its ROE is larger than those of 80% of all sample stocks. Similarly, a stock is assigned a score of 40 if its P/B ratio is lower than that of 40% of all sample stocks. Furthermore, we assume that the weights of the two stocks are 1/2 and 1/2. Then, the weighted scores are  $(1/2)*80 + (1/2)*40 = 60$ .

#### Definitions of performance indicators

Portfolio performance is evaluated through three categories: returns, risks, and liquidity. This paper adopts these three performance indicators as shown below:

#### Excess rate of return $\alpha$

Excess rate of return is estimated by the following regression equation:

$$R - R_f = \alpha + \beta (R_m - R_f) \quad (6)$$

where  $R_f$  = risk-free rate of return,  $R_m$  = market return,  $R$  = return of the portfolio. The equation denotes a positive excess return from the portfolio if  $\alpha > 0$ .

#### Systematic risk $\beta$

Systematic risk  $\beta$  can be estimated using Eq. (6). The higher the coefficient  $\beta$ , the higher the systematic risk. Portfolio volatility is higher than overall market if  $\beta > 1$ .

#### Stock market value in the portfolio

The larger the total market value of corporate stocks, the larger the trading volume of the stocks. The median of total market values of the portfolio's corporate stocks is chosen as a proxy variable in assessing the portfolio's liquidity.

#### Data set partition

A relationship exists between the weights of stock-picking concepts and each investment performance indicator. For instance, the qualitative relationship of stocks indicates that the returns of low-priced stocks with a small P/B ratio are usually higher than those of high-priced stocks with a large P/B ratio are. However, stocks' quantitative relationship can vary and change over time and thus can be highly volatile. Previous data cannot be used to construct models for selecting stocks for future investment when the quantitative relationship between the weights of stock-picking concepts and each investment performance indicator indicates high volatility. Therefore, it is necessary to consider the

time factor in order to explore whether the quantitative relationship is stable when we separate the data into two types, namely, in-sample and out-of-sample.

This study covers 19 years divided into two periods: the modeling and testing periods. Investment performances obtained through backtesting the weights of stock-picking concepts during the modeling period 1997–2008 and testing period 2009–2015 comprise the in-sample and out-of-sample data, respectively.

### **Experimental design**

The simplex-centroid design's q-type composition is expected to have  $2^q - 1$  experimental mixes. This study has five components, that is,  $2^5 - 1 = 31$  experimental mixes. Thus, 31 combinations of stock-picking concept weights are obtained. Each factor can be set up with six levels of weighting percentages, including 1, 1/2, 1/3, 1/4, 1/5, and 0 as exhibited on the left-hand side of Table 1.

### **Experimental implementation**

The following steps are used to obtain performance indicators through the 31 combinations of stock-picking concept weights proposed in this study.

#### ***Establishing a monthly database of corporate stocks***

We collect information on the P/B ratio, ROE, monthly rate of return, market values, and 250-day  $\beta$  of stocks between 1997 and 2015 from the Taiwan Stock Exchange and over-the-counter (OTC) markets.

#### ***Establishing monthly investment portfolios and calculating their performances***

The holding duration of investment portfolios is 1 month. We establish investment portfolios with the top 10% weighted-scoring stocks at the end of each month according to the 31 combinations of stock-picking concept weights in Table 1. We also calculate the median market value of the stocks in the portfolios and their following monthly rate of return of the portfolios.

#### ***Measuring the overall performances of each mixture design***

We calculate the monthly excess rate of return  $\alpha$  and systematic risk  $\beta$  using Eq. (6) with the portfolios' and market monthly rates of return. We also calculate the mean using the median market value of corporate stocks of each monthly portfolio.

The right-hand side of Table 1 presents the results of experimental implementation.

### **Model building and verification**

#### **Constructing the performance prediction model**

The dependent variables of the regression model are the three performance indicators, namely, excess rate of return  $\alpha$ , systematic risk  $\beta$ , and market value of the stocks in the portfolio. Hence, there are three regression models. Furthermore, the experimental results in Table 1 show that the distribution of medians of the market values deviates from the normal distribution. Therefore, natural-logs of market values were used to address the issue.

**Table 1 The mixture designs of stock-picking concepts and their experiment results**

No	Stock-picking concept weight				Performances of portfolio						
					Modeling period (1997–2008)			Testing period (2009–2015)			
	Small PBR	Large ROE	Large monthly return	Small beta ( $\beta$ )	Large total market value	Monthly excess rate of return (a) %	Monthly systematic risk ( $\beta$ )	Market value (100 million NTD)	Monthly excess rate of return (a) %	Monthly systematic risk ( $\beta$ )	Market value (100 million NTD)
1	1	0	0	0	0	1.100	0.937	27.9	0.864	1.008	22.6
2	0	1	0	0	0	1.070	0.891	54.6	1.094	0.876	71.5
3	0	0	1	0	0	0.480	1.144	55.7	0.619	0.976	53.0
4	0	0	0	1	0	0.720	0.834	27.7	1.069	0.610	25.5
5	0	0	0	0	1	0.312	0.562	639.1	0.092	0.848	620.2
6	1/2	1/2	0	0	0	1.250	0.864	32.8	1.382	0.908	35.2
7	1/2	0	1/2	0	0	1.097	0.947	37.7	0.875	0.990	30.6
8	0	1/2	1/2	0	0	1.018	1.076	45.2	1.263	0.900	70.1
9	1/2	0	0	1/2	0	1.200	0.713	13.3	1.162	0.757	21.8
10	0	1/2	0	1/2	0	1.306	0.767	58.6	1.582	0.738	48.9
11	0	0	1/2	1/2	0	0.950	0.945	33.1	1.193	0.721	34.8
12	1/2	0	0	0	1/2	0.545	0.762	159.2	0.281	0.922	148.4
13	0	1/2	0	0	1/2	0.712	0.764	368.7	0.595	0.825	311.1
14	0	0	1/2	0	1/2	0.409	0.772	333.6	0.434	0.887	265.1
15	0	0	0	1/2	1/2	0.724	0.798	186.8	0.614	0.677	141.2
16	1/3	1/3	1/3	0	0	1.480	1.012	39.6	1.620	0.918	37.0
17	1/3	1/3	0	1/3	0	1.450	0.857	20.3	1.575	0.776	24.8
18	1/3	0	1/3	1/3	0	1.235	0.711	30.9	1.216	0.840	19.9
19	0	1/3	1/3	1/3	0	1.250	0.799	45.2	1.544	0.780	50.4
20	1/3	1/3	0	0	1/3	1.030	0.717	188.7	0.714	0.857	181.3
21	1/3	0	1/3	0	1/3	0.608	0.958	156.0	0.476	0.921	137.0
22	0	1/3	1/3	0	1/3	0.692	0.721	273.1	0.922	0.837	204.4

**Table 1 (continued)**

No	Stock-picking concept weight				Performances of portfolio							
					Modeling period (1997–2008)				Testing period (2009–2015)			
					Small PBR	Large ROE	Large monthly return	Small beta ( $\beta$ )	Large total market value	Monthly excess rate of return ( $\alpha$ ) %	Monthly systematic risk ( $\beta$ )	Market value (100 million NTD)
23	1/3	0	0	1/3	1/3	0.934	0.841	119.1	0.627	0.747	93.7	
24	0	1/3	0	1/3	1/3	1.033	0.447	151.4	1.102	0.711	135.6	
25	0	0	1/3	1/3	1/3	0.593	0.758	102.5	0.890	0.714	130.3	
26	1/4	1/4	1/4	1/4	0	1.410	0.547	29.4	1.576	0.775	29.4	
27	1/4	1/4	1/4	0	1/4	0.931	1.021	125.2	0.978	0.839	169.0	
28	1/4	1/4	0	1/4	1/4	1.093	0.532	107.8	1.133	0.705	117.9	
29	1/4	0	1/4	1/4	1/4	0.888	0.577	54.1	0.813	0.782	97.5	
30	0	1/4	1/4	1/4	1/4	0.889	0.605	164.0	1.236	0.747	130.3	
31	1/5	1/5	1/5	1/5	1/5	1.034	0.973	79.8	1.293	0.724	108.9	

The independent variables of the regression model are the weights of the five stock-picking concepts. Regression analysis was conducted by way of polynomial regression in Eq. (1). Effects of the higher-order terms can be ignored because they are extremely small. This study selects the first, second, and third terms only. Given that the stock-picking concepts have five weights, each regression equation has five linear terms, 10 two-factor interaction terms, and 10 three-factor interaction terms, totaling 25 regression coefficients. A stepwise regression was adopted to eliminate certain insignificant terms.

Table 2 summarizes the regression coefficients and their t-statistics and significance from the models based on the three-order multivariable polynomial stepwise regression analysis. We use the regression coefficients to identify the impacts of the independent variables and their relationship with the dependent variables based on the following rules (Myers and Montgomery 2008; Montgomery 2012):

### 1 Coefficients of the linear terms

Linear terms have positive effects if their coefficient is larger than the average coefficient of linear terms, and vice versa.

**Table 2** Regression coefficients of regression models of portfolio performances

Term of the regression model	Monthly excess rate of return ( $\alpha$ ) of the portfolio			Monthly systematic risk ( $\beta$ ) of the portfolio			The mean of the median of the market value of the stocks in the portfolio		
	Coef	t-stat	Significance	Coef	t-stat	Significance	Coef	t-stat	Significance
PBR	1.088	*	*	0.900	*	*	3.348	*	*
ROE	1.077	*	*	0.874	*	*	3.980	*	*
R	0.479	*	*	1.121	*	*	3.908	*	*
beta	0.706	*	*	0.777	*	*	3.509	*	*
MV	0.324	*	*	0.588	*	*	6.616	*	*
PBR * ROE	0.782	4.100	0.001	—	—	—	—	—	—
PBR * R	1.318	6.890	0.000	—	—	—	—	—	—
PBR * beta	1.387	8.270	0.000	—	—	—	−3.494	−4.170	0.000
PBR * MV	−0.552	−2.920	0.012	—	—	—	—	—	—
ROE * R	1.021	5.400	0.000	—	—	—	—	—	—
ROE * beta	1.787	10.660	0.000	—	—	—	—	—	—
ROE * MV	—	—	—	—	—	—	3.370	4.220	0.000
R * beta	1.486	8.340	0.000	—	—	—	—	—	—
R * MV	—	—	—	—	—	—	2.613	3.110	0.005
beta * MV	0.922	5.170	0.000	—	—	—	—	—	—
PBR * ROE * R	5.994	4.660	0.000	—	—	—	—	—	—
PBR * ROE * beta	—	—	—	—	—	—	—	—	—
PBR * ROE * MV	3.615	3.010	0.010	—	—	—	—	—	—
PBR * R * beta	—	—	—	−8.020	−2.870	0.009	—	—	—
PBR * beta * MV	—	—	—	—	—	—	14.811	2.740	0.012
PBR * R * MV	−2.922	−2.420	0.031	—	—	—	—	—	—
ROE * R * beta	—	—	—	−4.223	−1.510	0.146	—	—	—
ROE * R * MV	−1.912	−1.670	0.118	—	—	—	—	—	—
ROE * beta * MV	—	—	—	−8.288	−2.960	0.007	—	—	—
R * beta * MV	−4.136	−3.430	0.004	—	—	—	−9.468	−1.750	0.095

Notation: PBR = Concept of small PBR; ROE = Concept of large ROE; R = Concept of large monthly return; beta = Concept of small beta; MV = Concept of large total market value

## 2 Coefficients of the quadratic term

The regression estimate between two independent variables is a convex function if the coefficient of the quadratic term is larger than zero. Otherwise, it is a concave function if the coefficient of the quadratic term is less than zero.

### *Monthly excess rate of return*

Table 2 shows that the two stock-picking concepts, namely, small P/B ratio and large ROE, positively impact monthly excess rate of return. In contrast, the other three stock-picking concepts, particularly large market capitalization concept, have negative impacts. These findings indicate that low-priced stocks tend to have higher return than high-priced stocks do. Similarly, profitable corporate stocks have higher return than less profitable ones do. These two long-term stock-picking concepts remain solid.

The concepts of large monthly return and small  $\beta$  negatively impact monthly excess rate of return. However, distinctively positive relationships exist among the four stock-picking concepts, namely, small P/B ratio, large ROE, large monthly return (R), and small  $\beta$  (beta) at the quadratic interactions terms, including P/B\*ROE, P/B\*R, P/B\*beta, ROE\*R, ROE\*beta, R\*beta, and beta\*MV. They are all significant at the level of 0.001. These results indicate the stock-picking concepts' synergy effects and ability to enhance returns. The effect is particularly outstanding in the case of ROE\*beta and signifies that the return of corporate stocks with a large ROE and small  $\beta$  is relatively stable relative to the return of corporate stocks only with a large ROE because of the attribution of corporate stock's small  $\beta$ . This finding also implies that low volatility can help to sustain profitability, then to obtain high returns.

Additionally, the monthly excess rate of return model has five significant coefficients of cubic terms. The first two, PBR\*ROE\*R and PBR\*ROE\*MV, are positive and share the characteristics of stock-picking concepts, namely, small P/B ratio and large ROE. The last three, PBR\*R\*MV, ROE\*R\*MV, and R\*beta\*MV, are negative and share the common characteristics of a stock-picking concept, that is, large total market value.

### *Monthly systematic risk $\beta$*

Two stock-picking concepts, namely, small  $\beta$  and large total market value, negatively impact monthly systematic risk  $\beta$ . The other three stock-picking concepts, namely, small P/B ratio, large ROE, and large monthly rate of return positively impact monthly systematic risk  $\beta$ . This finding indicates that the systematic risk of a portfolio, including corporate stocks with a small  $\beta$  and large total market value in the past, is relatively small. In contrast, the coefficients of the three cubic terms, P/B\*R\*beta, ROE\*R\*beta, and ROE\*beta\*MV, are significantly less than zero, whereas none of the coefficients of the quadratic terms are significant. The above three cubic terms share the common characteristics of a stock-picking concept, that is, small  $\beta$ . These results imply that the small  $\beta$  is the most important concept to lower portfolios' systematic risk, and the systematic risk can be reduced by considering additional stock-picking concepts.

### *Stock market value median of the portfolio*

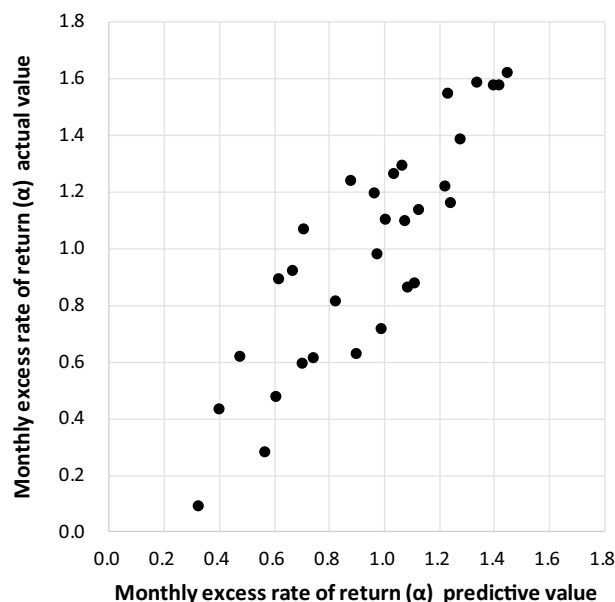
The concept of large total market value positively impacts the median market value of corporate stocks of the portfolio, whereas the other four stock-picking concepts have

negative impacts. Although linear terms of these four concepts have negative impacts, several of these concepts still have positive interactions with the concept of the large total market value and therefore can increase the median of market values of stocks in the portfolio. The most distinguished cases are  $ROE \cdot MV$  and  $R \cdot MV$ . However, the interaction term  $PBR \cdot \beta$  negatively impacts the median market value of stocks in the portfolio.

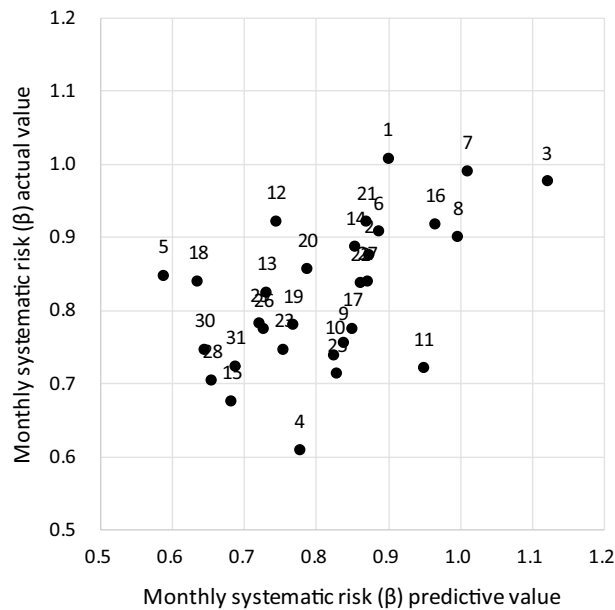
#### Out-of-sample prediction power of the regression models

The scatter diagrams in Figs. 6, 7 and 8 are drawn from the predicted values and actual values of the testing period data (out-of-sample) for the three regression models of performance indicators. The above data are produced from the 31 combinations of weights of stock-picking concepts through a mixture experimental design. These data help verify whether the prediction model based on the modeling period (1997–2008) data can also be applied to predict the performances during the testing period (2009–2015). The prediction model's out-of-sample prediction effects indicate that the mean of the median stock market value of the portfolio has the best accuracy, followed by monthly excess rate of return  $\alpha$ , whereas monthly  $\beta$  has the worst performance.

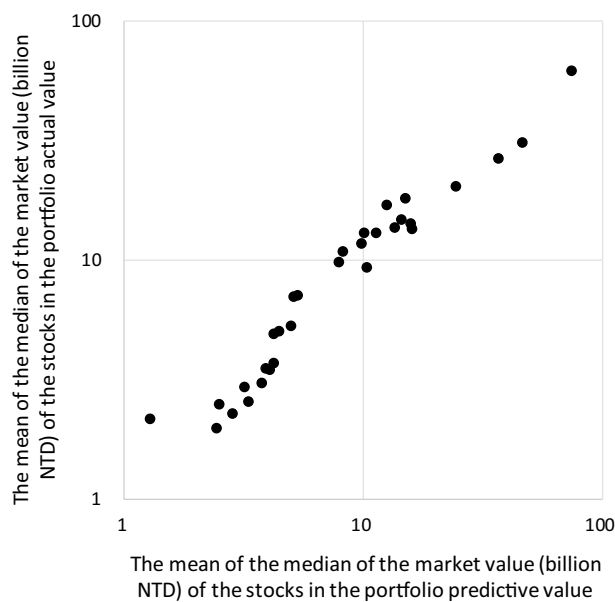
A few actual values of monthly  $\beta$  of the testing period data (out-of-sample) deviate from the predicted values of the modeling period data (in-sample). Figure 7 illustrates the code of the mixture design of the data to investigate further as to which data show a large deviation. Table 3 also presents their combinations of weights and indicates that large data deviation results from employing only one or two stock-picking concepts. Three of the five data through single stock-picking concepts, that is, large monthly rate of return, small  $\beta$ , and large market value, have a large deviation. Two of the 10 mixtures through two stock-picking concepts, and one of the 10 mixtures



**Fig. 6** Scatter diagram of the predicted values with regression models and the actual values of the testing period data (out-of-sample): monthly excess rate of return ( $\alpha$ )



**Fig. 7** Scatter diagram of the predicted values with regression models and the actual values of the testing period data (out-of-sample): monthly systematic risk ( $\beta$ )



**Fig. 8** Scatter diagram of the predicted values with regression models and the actual values of the testing period data (out-of-sample): The mean of the median of the market value (billion NT dollar) of the stocks in the portfolio

through three stock-picking concepts have large deviations. None of the mixtures through four or more stock-picking concepts has a large deviation. Thus, we may conclude that employing several stock-picking concepts can stabilize the relationship between the weights of stock-picking concepts and monthly systematic risk  $\beta$ .

**Table 3 Out-of-sample with large predictive deviation on the monthly systematic risk  $\beta$** 

No	Concept of small PBR	Concept of large ROE	Concept of large monthly return	Concept of small beta ( $\beta$ )	Concept of large market capitalization
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1
11	0	0	1/2	1/2	0
12	1/2	0	0	0	1/2
18	1/3	0	1/3	1/3	0

**Table 4 Adjusted coefficients of determination of the in-sample and out-of-sample**

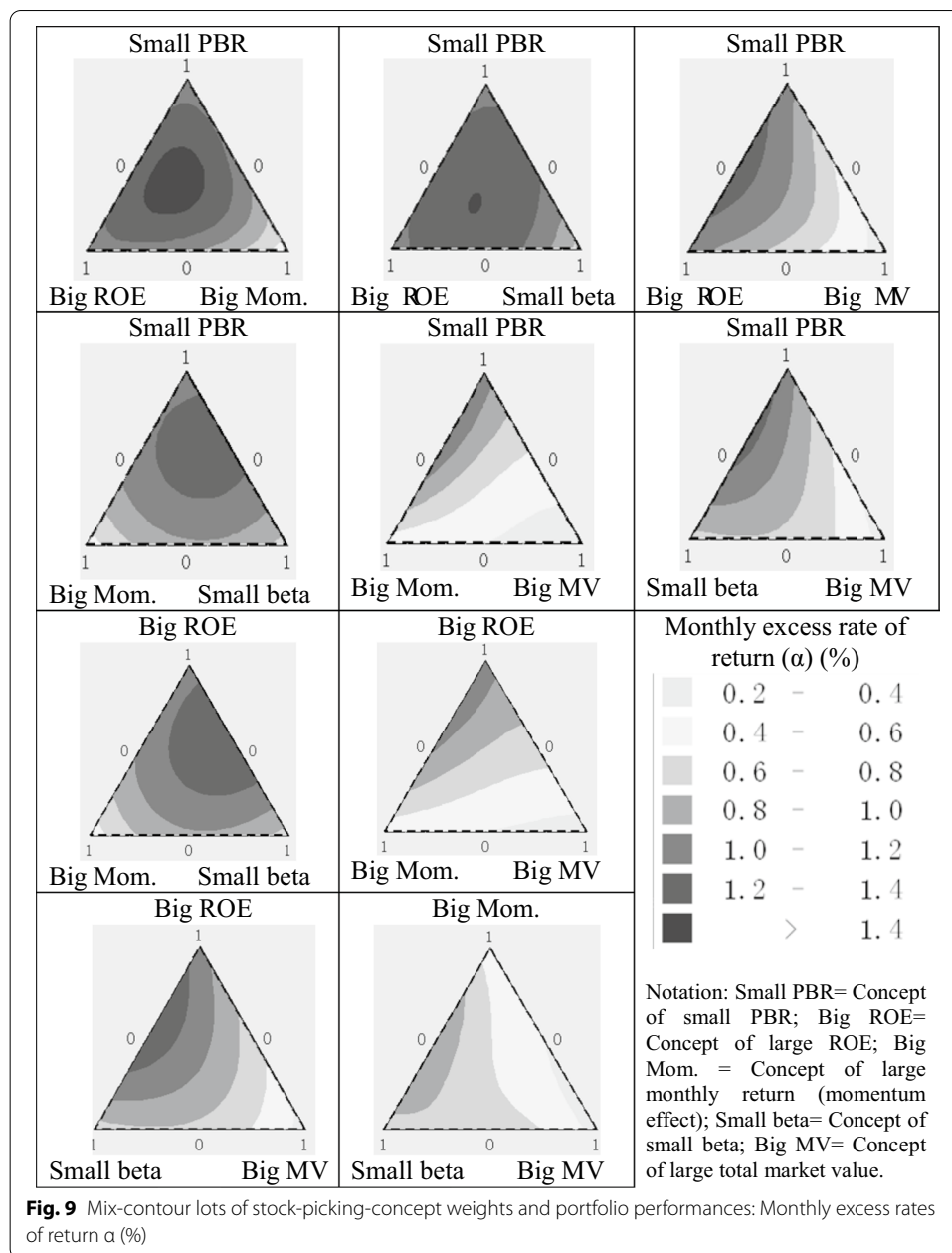
Period	Monthly excess rate of return ( $\alpha$ ) (%)	Monthly systematic risk ( $\beta$ ) (%)	Market capitalization
Performances of the Portfolio between 1997 and 2008 (in sample)	98.1	56.9	95.4
Performances of the Portfolio between 2009 and 2015 (out of sample)	76.4	31.9	94.2

Table 4 exhibits a comparison of adjusted coefficients of determination of the prediction model in the modeling period (in-sample) and testing period (out-of-sample). We found that the explanatory powers of the prediction model in the out-of-sample are lower than those in the in-sample are. Although the prediction model's explanatory power for monthly excess rate of return  $\alpha$  in the out-of-sample (76.4%) is lower than that in the in-sample (98.1%), its coefficient of determination maintains a high level. The prediction model's explanatory power for monthly systematic risk  $\beta$  in the out-of-sample (31.9%) is much lower than that in the in-sample (56.9%).

#### Visualization of the regression model

The mix-contour plots exhibit each prediction model to investigate the interactions of the weight of each stock-picking concept. The mix-contour plot in Fig. 9 is a regular triangle chart, whose apexes, sides, and interior are single-, two-, and three-component mixes, respectively. The midpoint of each side is a two-component mix (1/2, 1/2), whereas the centroid of the triangle is a three-component mix (1/3, 1/3, 1/3). The contour lines of the dependent variable (response) in the triangle are employed to visualize the impacts of each component on response.

Only three components can be shown in a triangle's mix-contour plot; hence, we select three of the five components and assume the other two components as zero to construct a mix-contour plot. If we select three of the five components each time, 10 combinations are generated, resulting in 10 mix-contour plots. Therefore, we produced 10 mix-contour plots from the five stock-picking concepts (components) adopted in this paper. Then, we investigate the mix-contour plot of each performance indicator (response) below.



### Excess rate of return

Responses from the midpoints of the sides of “small P/B–large ROE,” “small P/B–large momentum,” and “large ROE–large momentum,” are significantly higher than their two apexes according to the first mix-contour plot on the upper-left-hand side of Fig. 9. Therefore, the three sets of two-component mix have significantly positive interactions. Their regression coefficients are all statistically significant (5% threshold value) as shown in Table 2. Thus, the implication of the first mix-contour plot in Fig. 9 confirms the results in Table 2.

The second mix-contour plot of “small P/B–large ROE–small beta” denotes that the three sets of two-component mix have significantly positive interactions. These results are consistent with those in Table 2.

The third mix-contour plot of “small P/B–large ROE–large market value” signifies that the two sides have monotonous responses. The response increases if the side moves from the apex of large market value closer to either of the two other apexes, namely, the small P/B ratio and large ROE. These findings indicate that the two sets of two-component mix, namely, “small P/B–large market value” and “large ROE–large market value,” have no interactions, which are consistent with the results in Table 2. The same explorations can also be applied to the other plots.

#### **Monthly systematic risk $\beta$**

The first three charts on the upper side of Fig. 10 are “small P/B–large ROE–large momentum,” “small P/B–large ROE–small beta,” and “small P/B–large ROE–large market value” mix-contour plots. Their two-component mixes do not interact because the responses of each side are all monotonous.

The fourth, seventh, and ninth charts are the mix-contour plots of “small P/B–large momentum–small beta,” “large ROE–large momentum–small beta,” and “large ROE–small beta–large market value,” respectively. They are commonly characterized by the smallest responses of the triangle’s midpoints. In other words, their three-component mixes have significantly negative interactions. Moreover, their regression coefficients of the cubic terms are all negative values as shown in Table 2. Thus, these mix-contour plots match the results in Table 2.

#### **Total market capitalization**

The second mix-contour plot on the upper-left-hand side of Fig. 11 illustrates that the response from the midpoints of the sides of “small P/B–small beta” is significantly lower than that of the two apexes. Table 2 results show that this two-component mix has statistically significantly negative regression coefficient (5% threshold value), which is consistent with the mix-contour plot.

#### **Weight optimization and validation**

The most important benefit of a mixture experimental design is that it can provide an optimal composition of the mixture. Hence, this study can provide the optimal combination of weights of stock-picking concepts as follows.

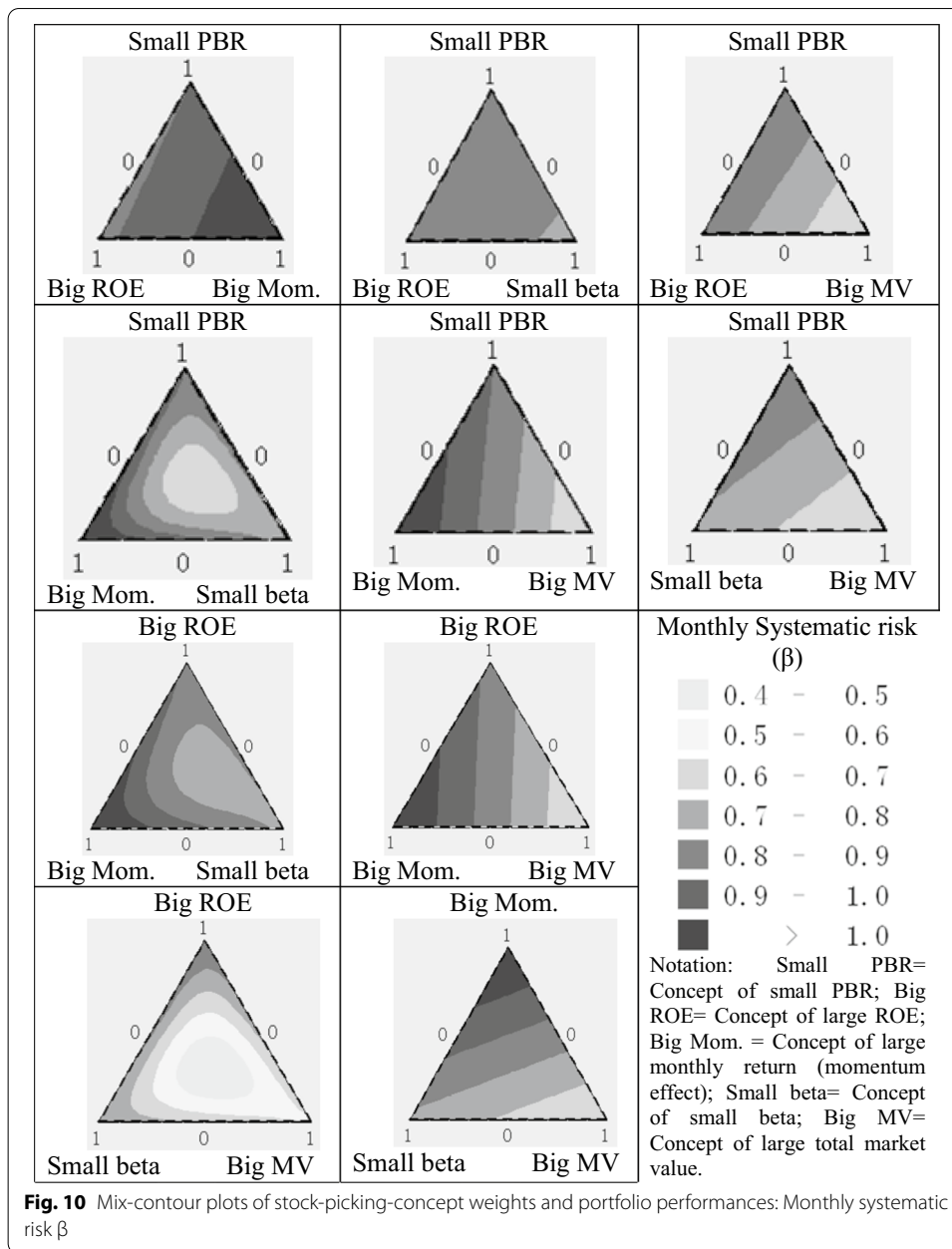
To find optimal weights,  $W$ , perform the following operations:

$$\text{Maximize } \alpha = f_{\alpha}(W) \quad (7)$$

Subjected to

$$\beta = f_{\beta}(W) \leq \text{specified upper bound} \quad (8)$$

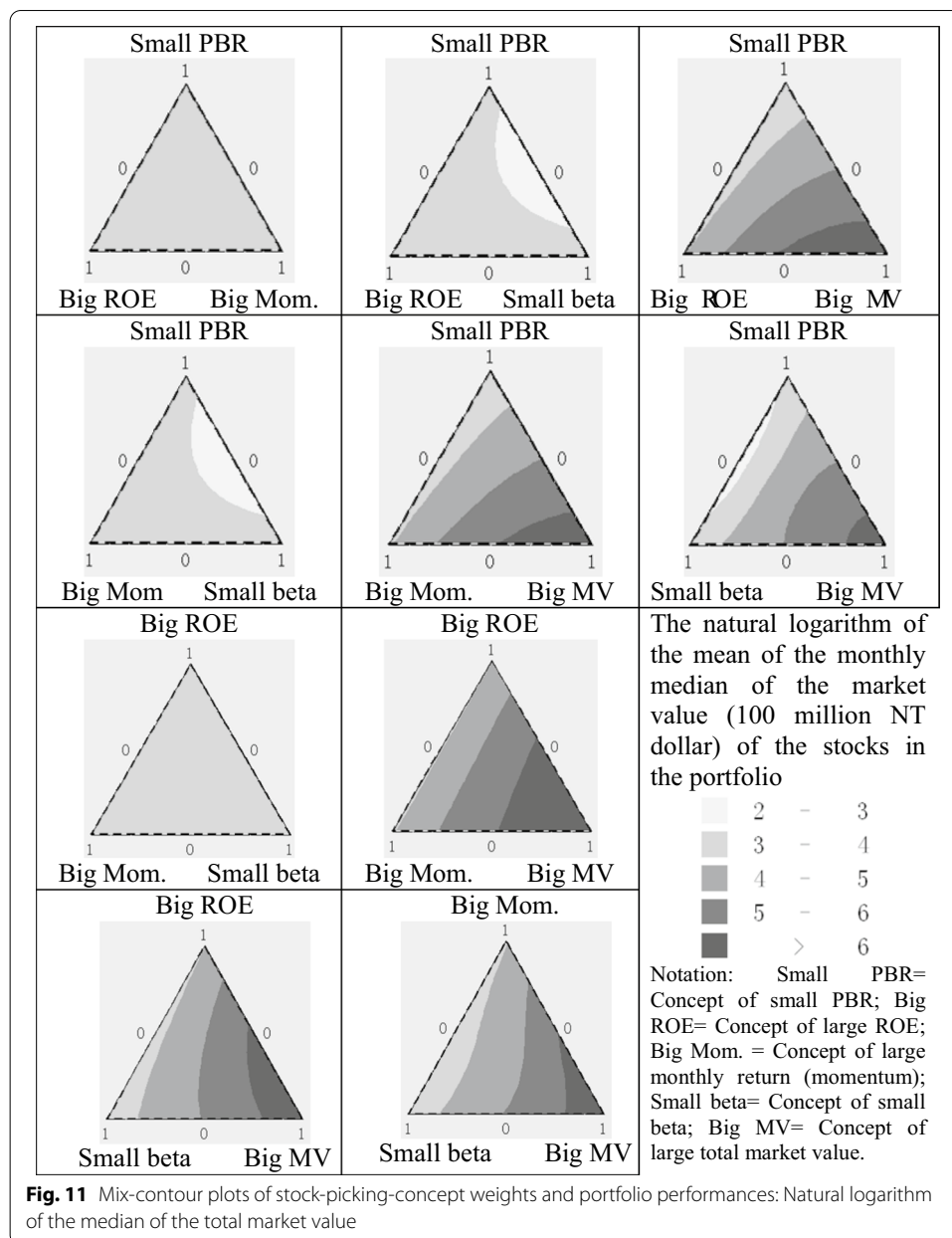
$$MV = f_{MV}(W) \geq \text{specified lower bound} \quad (9)$$



$$\sum_{i=1}^5 w_i = 1 \quad (10)$$

where  $\alpha = f_{\alpha}(W)$  = monthly excess rate of return,  $\beta = f_{\beta}(W)$  = monthly systematic risk, and  $MV = f_{MV}(W)$  = market value of the portfolio.

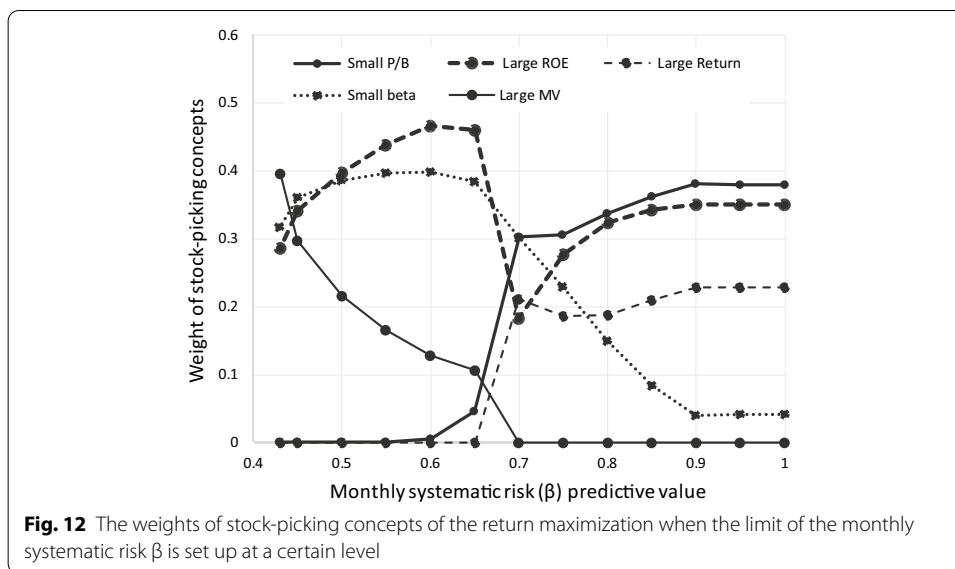
The above optimization model is a simple classical nonlinear programming problem, which can be solved using classical nonlinear programming algorithms. We used the generalized reduced gradient (GRG) algorithm to solve the optimization models. The details of the algorithm can be found in the literature (Nocedal and Wright 1999).



We can use the above model to determine the optimal combination of weights of stock-picking concepts. By doing this, we can maximize excess rate of return and limit the portfolio's systematic risk and market value to satisfy upper and lower bounds. Then, we apply the optimal combination of weights of stock-picking concepts to form a portfolio with the highest-scoring decile stocks.

#### Rate of Return Maximization with Risk Limitation

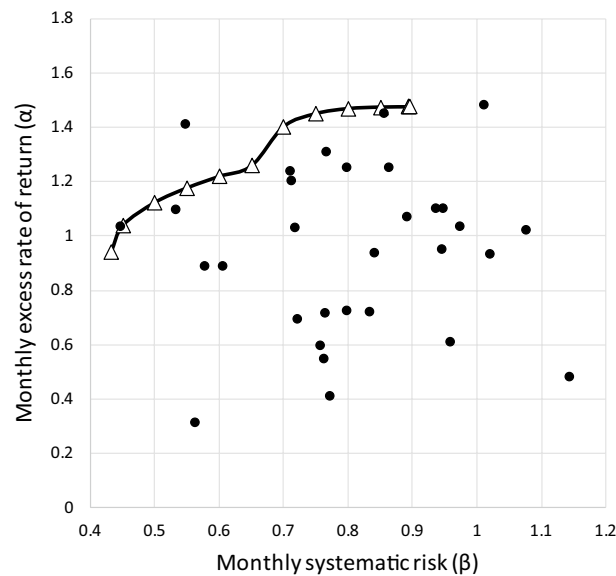
Figure 12 exhibits the weights of stock-picking concepts for maximizing the monthly excess rate of return  $\alpha$  by limiting the monthly systematic risk  $\beta$  to less than 1, 0.95, 0.9, ..., 0.55. Implications of the results of Fig. 12 include the following:



- (1) When the limit of the systematic risk  $\beta$  is set at a loose level ( $\beta > 0.9$ ), the weights of the small P/B ratio, large ROE, large momentum, and small beta are 38%, 35%, 23%, and 4%, respectively.
- (2) When the limit of systematic risk  $\beta$  is set at a relatively loose level ( $0.7 < \beta < 0.9$ ), the weights of the small P/B ratio, large ROE, and large momentum decrease, whereas that of the small beta increases.
- (3) When the limit of the systematic risk  $\beta$  is set at the middle level ( $0.6 < \beta < 0.7$ ), the weights of the small P/B ratio and large momentum drop sharply, whereas those of large ROE and small beta increase. Besides, the stock-picking concept of large market value becomes important.
- (4) When the limit of the systematic risk  $\beta$  is set at a strict level ( $\beta < 0.6$ ), the weights of large ROE and small beta decrease, whereas those of small P/B ratio and large momentum become zero. Then, the weight of large total market value becomes the most important. However, the optimal combination of weights of stock-picking concepts is not available if the limit of the systematic risk is set at a further stricter level lower than 0.4.

#### Validation in modeling period

Figure 13 illustrates the portfolio performances of the 31 combinations of weights of stock-picking concepts during the modeling period (1997–2008) in round, black spots. The upper-left-hand-side curve in Fig. 13 is the risk–return relationship curve comprising the predicted values of the prediction model of the optimal weights. This risk–return curve is drawn from the optimization model and close to the edge of the upper-left-hand-side area of the portfolio performances of the 31 combinations of weights of the mixture experimental design, forming a risk–return efficient frontier. This curve has 11 spots, which are estimated results when the limits of the monthly systematic risks  $\beta$  are set at the level of 0.9, 0.85, 0.8, ..., 0.45 and 0.4. The estimated



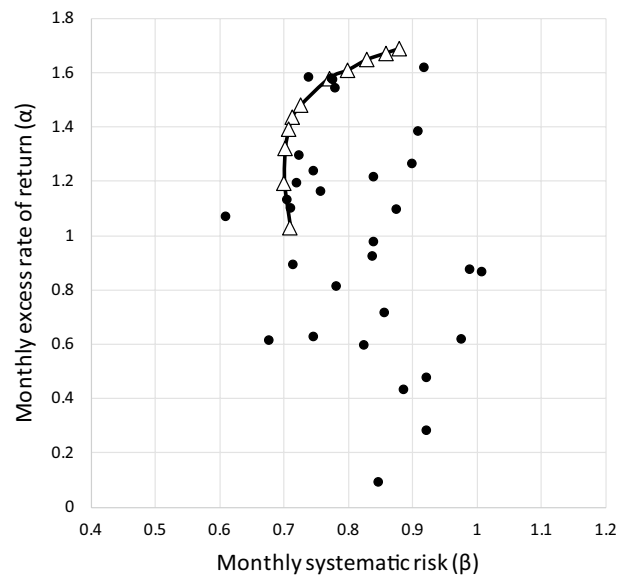
**Fig. 13** The risk-and-return relationship when the maximized return obtained by the monthly risk  $\beta$  limited at the specific level: modeling period (in-sample)

results are the same as the monthly systematic risk limit at the level between 0.9 and 1.0. Therefore, these performances generate overlapping spots, except for one of the portfolio performances of the 31 combinations of weights lying beyond the efficient frontier. This condition may be attributed to the prediction model's inaccuracy.

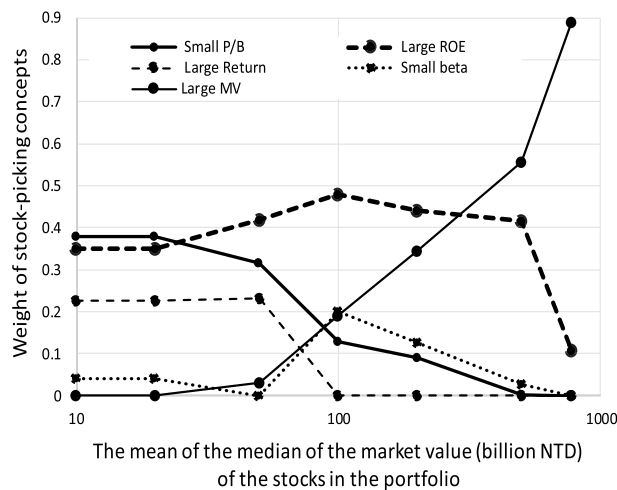
#### **Validation in testing period**

The optimization model is associated with the prediction model based on the performances during the modeling period (1997–2008). We further conduct backtesting on the testing period (2009–2015) using the above optimal weights generated by the optimization model to verify whether the optimal weights can also be applied to stock markets during this period. Figure 14 shows the portfolio performances of the 31 combinations of weights of stock-picking concepts during the testing period in round, black spots. The upper-left-hand-side curve depicts a risk–return relationship and is drawn through actual backtesting values of the optimal weights during the testing period. The 11 spots along this curve are the backtesting results when the limits of the monthly systematic risk  $\beta$  are set at the levels of 0.9, 0.85, 0.8, ..., 0.45, and 0.4. The risk–return curve is close to the edge of the upper-left-hand-side area of the portfolio performances of the 31 combinations of weights and forms a risk–return efficient frontier. Only two of the portfolio performances of the 31 combinations of weights lay beyond the efficient frontier. Therefore, we may conclude that the optimal weights have a good performance not only in the modeling period but also in the testing period.

In sum, Figs. 13 and 14 demonstrated that our approach could create a group of portfolios close to the risk–return efficient frontier not only during the modeling period (in-sample) but also during the testing period (out-of-sample).



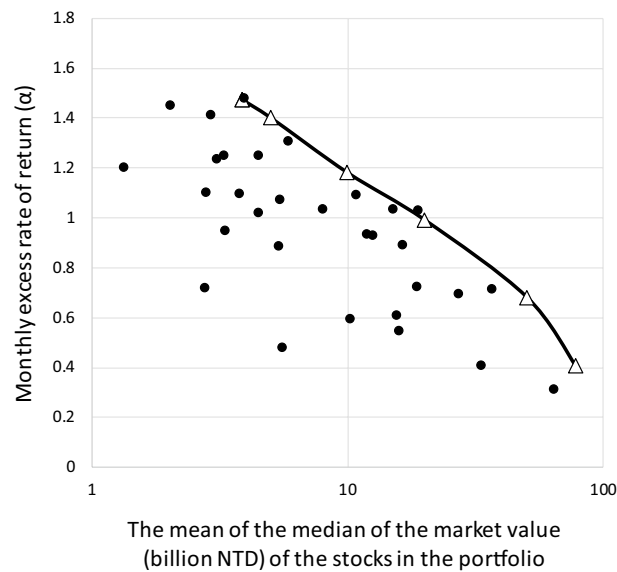
**Fig. 14** The risk-and-return relationship when the return maximization obtained by the monthly systematic risk  $\beta$  limited at the specific level: testing period (out-of-sample)



**Fig. 15** The weights of stock-picking concepts when the return rate maximization with the size requirement on the stock total market value

#### Rate of return maximization with market value limitation

Figure 15 displays the weights of stock-picking concepts for monthly excess rate of return  $\alpha$  maximization by limiting the market value to greater than 1, 2, 5, 10, 20, 50, and 100 billion NT dollars. The median market value of stocks in Taiwan Stock Exchange is approximately 3 billion NT dollars. The implications of the results in Fig. 15 include the following:

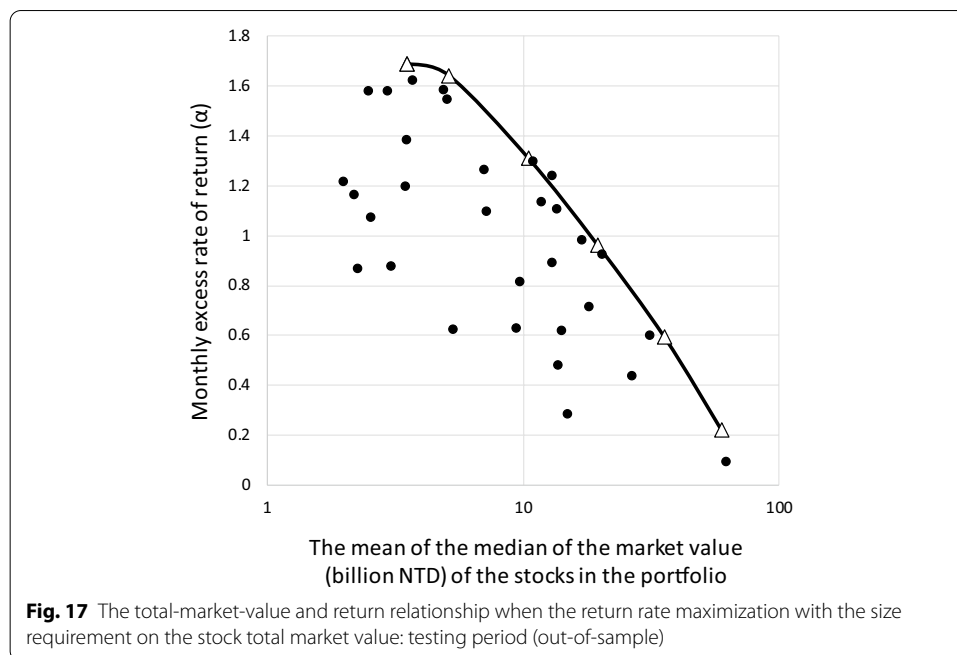


**Fig. 16** The total-market-value and return relationship when the return rate maximization with the size requirement on the stock total market value: modeling period (in-sample)

- (1) When the market value requirement is set at a lower level (<2 billion NT dollars), the weight of small P/B ratio, large ROE, large momentum, and small beta are 38%, 35%, 23%, 4%, respectively.
- (2) When the market value requirement is set at a normal level (2–5 billion NT dollars), the weights of large ROE, small P/B ratio, and large momentum increase, decrease, and remain unchanged, respectively.
- (3) When the market value requirement is set at a high level (5–10 billion NT dollars), the weights of large ROE, small P/B ratio, and large momentum increase, decrease, and becomes zero, respectively. Then, the concepts of small beta and large market value become more important.
- (4) When the market value requirement is set at an extremely high level (10–50 billion NT dollars), the weights of small P/B ratio and small beta gradually decrease to zero. Moreover, the weight of large ROE slightly becomes lower, and large market value stock-picking concept becomes the most important. No weight combination is available from the optimization model when the size requirement is larger than 100 billion NT dollars. The optimal combination of weights with the greatest market value of 78 billion NT dollars is that of large market value at and large ROE concepts at 90% and 10%, respectively.

#### **Validation in modeling period**

The 31 round, black spots shown in Fig. 16 are the portfolio performances of the 31 combinations of weights of stock-picking concepts during the modeling period. The curve on the upper-right-hand side comprises the prediction model's predicted values of optimal weights drawn from the optimization model and depicts the relationship between market value and return. This curve is close to the edge of the upper-right-hand-side area of



the portfolio performances of the 31 combinations of weights of the mixture experimental design. The curve forms a market-value-and-return efficient frontier. The six spots along the curve are the estimated results when market value size requirements are 2, 5, 10, 20, 50, and 100 billion NT dollars. The prediction model is significantly accurate because no spot lies beyond the efficient frontier.

#### **Validation in testing period**

We further conduct backtesting on the testing period with the above optimal weights to verify whether the optimal weights can also be applied to the stock markets of the testing period. The round, black spots in Fig. 17 depict the portfolio performances of the 31 combinations of weights of the mixture experimental design during the testing period. The upper-right-hand-side curve depicts the relationship between market value and return, which comprises the actual backtesting values. The six spots along this curve are the backtesting results when market value requirements are set at 2, 5, 10, 20, 50, and 100 billion NT dollars. This curve is close to the upper-right-hand-side edge of round, black spots, which depicts market value and return performances of the portfolio of the 31 combinations of weights. The curve forms a market-value-and-return efficient frontier because no spot lies beyond the curve. This condition signifies that the optimal weights have a good performance not only in the modeling period but also in the testing period.

#### **Conclusion**

Considerable literature has revealed that the more factors are included, the higher the rate of return would be. Several studies adopted a weighted-scoring approach to construct a multifactor stock selection model. However, this method sets up weights

subjectively or uses a simple average and thus cannot effectively identify the connection between the weights of stock-picking concepts and portfolio performances, provide optimal weights of stock-picking concepts, and meet various investor preferences.

This study addresses these drawbacks by employing mixture experimental designs to collect the weights of stock-picking concepts and portfolio performance data and to construct performance prediction models based on the weights of stock-picking concepts. Moreover, we employed these performance prediction models and optimization techniques to determine the optimal combination of weights of stock-picking concepts.

The samples consist of all stocks listed in the Taiwan Stock Exchange. Backtesting is conducted on the 19 years between 1997 and 2015. The 1997–2008 and 2009–2015 periods are employed as the modeling period (in-sample) and testing period (out-of-sample), respectively. The results provide important implications for stock investment.

First, mixture experimental designs and multivariable polynomial regression can construct performance prediction models based on the data set from the training period. These models are accurate not only during the training period but also during the testing period.

Second, the methodology can discover significant interactions between the weights of stock-picking concepts. The ROE and beta significantly positively impact the portfolios' excess rate of return and hence can effectively increase portfolio's return.  $P/B \times R \times \text{beta}$ ,  $\text{ROE} \times R \times \text{beta}$ , and  $\text{ROE} \times \text{beta} \times \text{MV}$  significantly negatively impact portfolios' systematic risk  $\beta$ . Thus, they can effectively reduce portfolio's risk. Furthermore,  $\text{ROE} \times \text{MV}$  and  $R \times \text{MV}$  significantly positively impact portfolios' market value. Therefore, they can effectively increase the portfolio's liquidity.

Third, the optimization techniques can efficiently determine the optimal combination of weights of factors that can form stock portfolios with the best possible performance and can meet various investor preferences.

Thus, our methodology can resolve the three drawbacks of classical weighted-scoring approach.

#### Abbreviations

Beta: Systematic risk; MV: Market value; PBR: Price-to-book value ratio; ROE: Return on equity; R: Monthly return.

#### Authors' contributions

The first author conducted the project, wrote the paper and revised it. The second author checked writing and approved the final version. All authors read and approved the final manuscript.

#### Funding

We do not receive any financial assistance from any agency.

#### Availability of data and materials

The dataset on which the conclusions of the manuscript rely is a secondary data and it will be made available upon request.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup> Department of Civil Engineering, Tamkang University, 151 Yingzhuan Rd., Tamsui Dist., New Taipei City 25137, Taiwan.

<sup>2</sup> Department of International Business, Tamkang University, 151 Yingzhuan Rd., Tamsui Dist., New Taipei City 25137, Taiwan.

Received: 23 April 2020 Accepted: 4 October 2020

Published online: 23 November 2020

## References

- Asness CS, Moskowitz TJ, Pedersen LH (2013) Value and momentum everywhere. *J Finance* 68(3):929–985
- Banz RW (1981) The relationship between return and market value of common stocks. *J Financ Econ* 9(1):3–18
- Bondt W, Thaler R (1985) Does the stock market overreact? *J Finance* 40(3):793–805
- Dai J, Zhou J (2019) A novel quantitative stock selection model based on support vector regression. In: 2019 international conference on economic management and model engineering (ICEMME), IEEE, pp 437–445
- Daniel K, Mota L, Rottke S, Santos T (2020) The cross-section of risk and returns. *Rev Financ Stud* 33(5):1927–1979
- Duran-Vazquez R, Lorenzo-Valdes A, Castillo-Ramirez CE (2014) Effectiveness of corporate finance valuation methods: Piotroski score in an Ohlson model: the case of Mexico. *J Econ Finance Admin Sci* 19(37):104–107
- Fama EF, French KR (2012) Size, value, and momentum in international stock returns. *J Financ Econ* 105(3):457–472
- Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *Rev Financ Stud* 33(5):2223–2273
- Hart JV, Slagter E, Dijk DV (2003) Stock selection strategies in emerging markets. *J Empir Finance* 10(1–2):105–132
- Hart JV, Zwart G, Dijk DV (2005) The success of stock selection strategies in emerging markets: Is it risk or behavioral bias? *Emerg Markets Rev* 6(3):238–262
- Hong H, Stein JC (1999) A unified theory of under-reaction, momentum trading and overreaction in asset markets. *J Finance* 54(6):2143–2184
- Hong H, Lim T, Stein JC (2000) Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. *J Finance* 55(1):265–295
- Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. *J Finance* 48(1):65–91
- Jeong T, Kim K (2019) Effectiveness of F-SCORE on the loser following online portfolio strategy in the Korean value stocks portfolio. *Am J Theor Appl Bus* 5(1):1–13
- Kang J, Ding D (2006) Value and growth investing in Asian stock markets 1991–2002. *Res Finance* 22:113–139
- Kim S, Lee C (2014) Implementability of trading strategies based on accounting information: Piotroski (2000) revisited. *Eur Acc Rev* 23(4):553–558
- Kong D, Lin CP, Yeh IC, Chang W (2019) Building growth and value hybrid valuation model with errors-in-variables regression. *Appl Econ Lett* 26(5):370–386
- Mehta N, Pothula VK, Bhattacharyya R (2019) A value investment strategy that combines security selection and market timing signals. SSRN 3451859
- Mohanram S (2005) Separating winners from losers among low book-to-market stocks using financial statement analysis. *Rev Account Stud* 10(3):133–170
- Montgomery DC (2012) Design and analysis of experiments. Wiley, New York, pp 611–622
- Myers RH, Montgomery DC (2008) Response surface methodology. Wiley, New York
- Nocedal J, Wright SJ (1999) Numerical optimization. Springer, New York
- Noma M (2010) Value investing and financial statement analysis. *Hitotsubashi J Commer Manag* 44(1):29–46
- Piotroski JD (2000) Value investing: the use of historical financial statement information to separate winners from losers. *J Account Res* 38:1–41
- Qian EE, Hua RH, Sorensen EH (2007) Quantitative equity portfolio management: modern techniques and applications. Chapman and Hall/CRC, Boca Raton
- Rasekhschaffe KC, Jones RC (2019) Machine learning for stock selection. *Financ Anal J* 75(3):70–88
- Richardson S, Tuna I, Wysocki P (2010) Accounting anomalies and fundamental analysis: a review of recent research advances. *J Account Econ* 50(2):410–454
- Roko I, Gilli M (2008) Using economic and financial information for stock selection. *CMS* 5(4):317–335
- Rosenberg B, Reid K, Lanstein R (1985) Persuasive evidence of market inefficiency. *J Portf Manag* 11(3):9–17
- Shen KY, Tzeng GH (2015) Combined soft computing model for value stock selection based on fundamental analysis. *Appl Soft Comput* 37:142–155
- Tikkanen J, Äijö J (2018) Does the F-score improve the performance of different value investment strategies in Europe? *J Asset Manag* 19(7):495–506
- Wen F, Xu L, Ouyang G, Kou G (2019) Retail investor attention and stock price crash risk: Evidence from China. *Int Rev Financ Anal* 65:101376
- Wu X, Ye Q, Hong H, Li Y (2020) Stock selection model based on machine learning with wisdom of experts and crowds. *IEEE Intell Syst* 35(2):54–64
- Yeh IC, Hsu TK (2011) Growth value two-factor model. *J Asset Manag* 11(6):435–451
- Yeh IC, Hsu TK (2014) Exploring the dynamic model of the returns from value stocks and growth stocks using time series mining. *Expert Syst Appl* 41(17):7730–7743
- Yeh IC, Lien CH, Ting TM (2015) Building multi-factor stock selection models using balanced split regression trees with sorting normalization and hybrid variables. *Int J Foresight Innov Policy* 10(1):48–74
- Yu H, Chen R, Zhang G (2014) A SVM stock selection model within PCA. *Procedia Comput Sci* 31:406–412

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.