

Wu, Wenbo; Chen, Jiaqi; Xu, Liang; He, Qingyun; Tindall, Michael L.

## Article

# A statistical learning approach for stock selection in the Chinese stock market

Financial Innovation

## Provided in Cooperation with:

Springer Nature

*Suggested Citation:* Wu, Wenbo; Chen, Jiaqi; Xu, Liang; He, Qingyun; Tindall, Michael L. (2019) : A statistical learning approach for stock selection in the Chinese stock market, Financial Innovation, ISSN 2199-4730, Springer, Heidelberg, Vol. 5, Iss. 1, pp. 1-18, <https://doi.org/10.1186/s40854-019-0137-1>

This Version is available at:

<https://hdl.handle.net/10419/237165>

### 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/>

RESEARCH

Open Access



# A statistical learning approach for stock selection in the Chinese stock market

Wenbo Wu<sup>1</sup>, Jiaqi Chen<sup>2</sup>, Liang Xu<sup>3</sup>, Qingyun He<sup>3\*</sup> and Michael L. Tindall<sup>2</sup>

\* Correspondence:

1178712318@qq.com

<sup>3</sup>School of Business Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

Full list of author information is available at the end of the article

## Abstract

Forecasting stock returns is extremely challenging in general, and this task becomes even more difficult given the turbulent nature of the Chinese stock market. We address the stock selection process as a statistical learning problem and build cross-sectional forecast models to select individual stocks in the Shanghai Composite Index. Decile portfolios are formed according to rankings of the forecasted future cumulative returns. The equity market's neutral portfolio—formed by buying the top decile portfolio and selling short the bottom decile portfolio—exhibits superior performance to, and a low correlation with, the Shanghai Composite Index. To make our strategy more useful to practitioners, we evaluate the proposed stock selection strategy's performance by allowing only long positions, and by investing only in A-share stocks to incorporate the restrictions in the Chinese stock market. The long-only strategies still generate robust and superior performance compared to the Shanghai Composite Index. A close examination of the coefficients of the features provides more insights into the changes in market dynamics from period to period.

**Keywords:** Stock selection, Stock return prediction, Statistical learning, Lasso, Elastic net

## Introduction

China's annual GDP growth has averaged more than 9% since 2000. Meanwhile, the Chinese stock market has experienced substantial fluctuations during this period. The Shanghai Composite Index reached a historical high of over 6000 in October 2007, only to fall to approximately half that level at present. This disparity between rapidly growing GDP and lackluster stock market performance poses a dilemma for investors. How can one benefit from the burgeoning Chinese economy by investing in the Chinese stock market? As Rapach and Zhou (2013) have indicated, forecasting stock returns is extremely challenging in general; this task is even more difficult when it involves the turbulent Chinese stock market. Kang et al. (2002) studied contrarian and momentum strategies for the Chinese stock market in earlier years, while Li et al. (2017) examined the performance of trends following strategies in Chinese commodity futures markets. The recent introduction of advanced statistical approaches to portfolio management, such as machine learning, has been highly successful (e.g. Li and Hoi 2015; Shen and Wang 2017; Wu et al. 2018; Gu et al. 2019). This paper presents a statistical learning approach to select top-performing Chinese stocks that can potentially generate superior returns and significantly outperform the market.

Some researchers have attempted to identify the underlying drivers for Chinese stock returns. Wang and Xu (2004) discovered that the value factor does not explain the cross-sectional differences in the Chinese stock market due to the market's speculative nature and low-quality accounting information. On the other hand, the size factor does carry certain explanatory power. Kling and Gao (2008) explored the positive feedback process between Chinese stock share prices and institutional investors' sentiment. Further, Yuan et al. (2008) found that the equity ownership of mutual funds positively affects performance. These types of information can all contribute to the stock selection process. However, these methods benefit institutional rather than individual investors, as the latter cannot easily access this information.

Given the Chinese stock market's poor performance, an increasing number of institutions and investors are searching for stocks with high excess returns compared to the Shanghai Composite Index. Empirical finance often ranks stocks according to a risk factor over a look-back window, and long and short stock positions are subsequently taken according to this ranking. Literature has also proposed numerous factors (Fama and French 1992; Ang et al. 2006; Frazzini and Pedersen 2014), with Harvey et al. (2016) identifying more than 200 such factors. Factor-ranking procedures are simple and effective, but have some inherent problems. For example, a particular factor's risk premium may be unstable over time. Fama and French (1992) found that their value factor exhibits a positive risk premium over the long-term in the U.S. market; however, we discovered that the value premium—estimated as the annual return difference between the Russell 1000 Value Index and Russell 1000 Growth Index—has been negative in all five-year rolling windows from 2009 to 2015.

The simple factor-ranking approach can be extended to include multiple risk factors in the portfolio-formation process, but this creates another problem. If we follow a traditional approach, we would select stocks ranked in the top group for all risk factors. However, a high-dimensional problem occurs with many factors, in that the stocks would be scattered sparsely in an  $N$ -dimensional space, with  $N$  being the number of factors that we would incorporate. It may be impossible to find enough stocks in the top rankings of all factors to form a diversified portfolio. One could use an ad hoc method by applying a composite score to combine the factors into a ranking. Mohanram (2005) combined traditional fundamental factors to create an index-GSCORE. While a long-short strategy based on the GSCORE earns significant excess returns, this method lacks theoretical guidance about how to assign weights among factors, and subjective factors always influence the method's precision in forming a composite score. These problems make the method unstable in practice. As previously mentioned, both simple and multiple factor-ranking approaches cannot effectively construct a portfolio with the highest future return, and thus, we must consider other ways to improve performance.

Cross-sectional regression plays an important role in finance to explain variations in stock prices (Sharpe 1964; Fama and French 1992; Carhart 1997). For example, Fama and French (1992) indicate that the cross-section regression provides a good description of returns on portfolios formed based on size, BE/ME, and term-structure risk factors. This leads to the gradual development of a method for determining factors' weight through regressions. While some subsequent works (Hou et al. 2015; Fama and French 2015) attempted to discover more factors to explain stock returns using

cross-sectional regression models, the regressions' accuracy is undermined by the effects of multicollinearity and overfitting using a large number of factors.

In this paper, we propose an advanced statistical model, the elastic net (Zou and Hastie 2005), to resolve this issue when regressing stock returns over a large number of factors. Our method differs from most existing methods by building cross-sectional forecast models for stock returns, and selects stocks based on these models' predicted returns. Our approach is rooted in the Fama-French-Carhart model (Carhart 1997), hereafter "the FFC four-factor model," but greatly expands its scope to include more statistical factors. Given the low quality of accounting information in the Chinese stock market, we focus on forecasting future returns using only the statistical factors derived from historical stock prices. We handle this as a supervised statistical learning problem in which portfolios are formed according to the forecast returns' rankings. We find that the highest ranked portfolios generate robust and superior performance. Our forecast methods deviate from traditional econometric approaches and are more in line with the approach taken by Varian (2014), who argues that data analysis in statistics and econometrics can be divided into four categories: 1) prediction, 2) summarization, 3) estimation, and 4) hypothesis testing. The statistical learning methods that we employ fall into the first category as our goal is to identify the stocks with the highest future returns.

This paper contributes to existing literature in the following ways. First, the proposed statistical learning approaches effectively rank, and hence select, the top performing stocks relying on the predictability of the forecast model. This is different from traditional cross-sectional regression models (Fama and French 1992; Carhart 1997). Second, building a statistical learning model provides a data-adaptive guidance on how to combine information from different factors (which we treat as features) in the proposed forecast model. Third, the use of the elastic net estimator (Zou and Hastie 2005) improves the estimation precision (accuracy) of the forecast model when multicollinearity is present and also avoids the overfitting that occurs when a large number of features are considered. Another advantage of the proposed method is its interpretability and feature selection capability. Lastly, the proposed stock selection framework is flexible and can be modified by investors to include user-specific features. The remainder of our paper is organized as follows: Section 2 discusses our motivation and explains the proposed methodology. Subsequently, Section 3 presents the data and features used in this study. Section 4 presents the empirical results, while the final section concludes.

## Methodology

### Motivation

Instead of relying solely on factor rankings, we use the FFC four-factor model to explain how our stock selection framework relates to traditional studies on cross-sectional returns. A stock's excess return can be decomposed based on the FFC four-factor model into the excess return of four factors—the market return, value, size, and momentum—as in the following equation:

$$R_t^i = \alpha^i + \beta_{MKT}^i R_{MKT,t} + \beta_{HML}^i R_{HML,t} + \beta_{SMB}^i R_{SMB,t} + \beta_{UMD}^i R_{UMD,t} + \varepsilon_t^i, \quad (1)$$

where  $R_t^i$  is the monthly excess return of a particular stock  $i$  over the Treasury bill rate;  $R_{MKT,t}$  is the market index's monthly excess returns;  $R_{HML,t}$  is the monthly excess

return of a zero-investment portfolio that is long on high book-to-market (B/M) stocks and short on low B/M stocks;  $R_{SMB, t}$  is a zero-investment portfolio that is long on small capitalization (cap) stocks and short on big-cap stocks;  $R_{UMD, t}$  is a zero-cost portfolio that is long previous on the 12-month return winners (i.e., returns in the top 30%) and short on the previous 12-month loser stocks (i.e., returns in the bottom 30%); and  $\varepsilon_t^i$  is the unexplained variation of stock  $i$  at time  $t$ .

Following Haugen and Baker (1996), the FFC four-factor model can be modified into a cross-sectional forecast model:

$$E(R_{t+1}^i | \phi_t^i) = \alpha^i + \beta_{MKT}^i R_{MKT,t} + \beta_{HML}^i R_{HML,t} + \beta_{SMB}^i R_{SMB,t} + \beta_{UMD}^i R_{UMD,t} + \varepsilon_{t+1}^i, \tag{2}$$

where the conditional expectation  $E(R_{t+1}^i | \phi_t^i)$  is the forecast return of stock  $i$  at time  $t + 1$  based on  $\phi_t^i$ , the information pertaining to stock  $i$  at time  $t$ . In Equation (2),  $\phi_t^i$  is the information represented by stock-specific betas. This model can be estimated through either a cross-sectional regression or a Fama-MacBeth process (Fama and MacBeth 1973).

All four factors in the FFC four-factor model are market-related risk premiums. Despite its effectiveness in explaining cross-sectional returns, we suspect that these four factors may be insufficient to model emerging markets, such as the Chinese stock market. Therefore, we propose a more comprehensive feature information set  $\Omega_t^i$  to be used in the Equation (2) forecast model by including past returns for stocks as well as other quantitative factors for stock  $i$ . We aim to use statistical learning methods to build a flexible modeling platform that can combine the strengths of all factors without relying on a heuristic argument to define the weights in the composite score. This can potentially improve model performance and to build a general framework that can be easily modified to incorporate new user-specified risk factors.

### Proposed method

Our method can be perceived as the following supervised statistical learning problem; at each time point  $t$ , the cross-sectional forecast model can be written as:

$$\tilde{R}_{t+1,t+F}^i = g(\Omega_t^i) + \varepsilon_{t+1,t+F}^i, \tag{3}$$

where  $\tilde{R}_{t+1,t+F}^i = \prod_{j=1}^F (1 + R_{t+j}^i) - 1$  is the cumulative return for stock  $i$  from time  $t + 1$  to  $t + F$  and  $g(\Omega_t^i)$  is the linear or nonlinear forecast function using past information from stock  $i$  up to time  $t$ . The past information  $\Omega_t^i$  can be either the historical returns of stock  $i$ :  $R_t^i, R_{t-1}^i, \dots$  at time  $t, t - 1, \dots$  or other quantitative factors or characteristics:  $f_{1,t}^i, f_{2,t}^i, \dots$  describing stock  $i$  at time  $t$ , such as the historical market alpha. To avoid ambiguity in later discussion, we call both the return based information and characteristic factors in  $\Omega_t^i$  the features and denote them by  $X_{1,t}^i, X_{2,t}^i, \dots$  following the conventional notations in statistics. In the statistics language,  $\tilde{R}_{t+1,t+F}^i$  is usually referred as the target or response variable of model (3). Further, Section 3 provides a complete list of features used in this study. Once the forecast function  $g(\cdot)$  is estimated, the forecast future cumulative returns are denoted by the expectation  $E(\tilde{R}_{t+1,t+F}^i | \Omega_t^i) = g(\Omega_t^i)$ .

This paper uses a linear forecast function  $g(\cdot)$  for Equation (3), as the linear model's interpretability is useful when examining how each factor contributes to the final portfolio's composition. Hence, if we assume that our study includes a total of  $p$  features, the cross-sectional linear forecast model in Equation (3) at time  $t$  can be explicitly written as

$$\tilde{R}_{t+1,t+F}^i = g_t(\Omega_t^i) + \varepsilon_t^i = \beta_0^t + \sum_{j=1}^p \beta_j^t X_{j,t}^i + \varepsilon_{t+1,t+F}^i, i = 1, \dots, n; \tag{4}$$

where  $\varepsilon_{t+1,t+F}^i$  is the unexplained variation of stock returns at  $t$ .

We use a rolling window to build our forecast model and evaluate its performance following the approach used by Moskowitz et al. (2010). Assume a look-back period of  $B$  and a look-forward period of  $F$ . At each time  $t$ , we must estimate the loading  $\beta_j^t$  s in Equation (3) to forecast the future cumulative return  $\tilde{R}_{t+1,t+F}^i$  for the time  $t + 1$  to  $t + F$  of each stock  $i$ . To estimate  $\beta_j^t$  s at time  $t$ , we compute features  $\Omega_{t-F}^i = X_{1,t-F}^i, X_{2,t-F}^i, \dots, X_{p,t-F}^i$  based on the information of each stock  $i$  from  $t - B - F + 1$  to  $t - F$  and compute the response variable  $\tilde{R}_{t+1,t+F}^i$  as described above from  $t - F + 1$  to  $t$ . Then a cross-sectional linear regression using information from all stocks  $i = 1, 2, \dots, n$  is fitted to obtain the estimated loading  $\hat{\beta}_j^t$  s which leads to an estimated forecast model  $\hat{g}_t(\cdot)$ . The following section will provide a more detailed estimation procedure. We then recalculate features  $\Omega_t^i = X_{1,t}^i, X_{2,t}^i, \dots, X_{p,t}^i$  based on the information of each stock  $i$  from  $t - B + 1$  to  $t$  and obtain the predicted future cumulative return  $\hat{R}_{t+1,t+F}^i = \hat{g}_t(\Omega_t^i)$  for each stock  $i$  in the  $F$  time periods following time  $t$ .

Based on the forecast  $\hat{R}_{t+1,t+F}^i$ , we rank the stocks into 10 equally sized groups to form 10 equally weighted portfolios (i.e., the decile portfolios) and hold these 10 portfolios for the next  $F$  months, after which we rebalance.<sup>1</sup> When rebalancing at the end of  $F$  months, we retrain the forecast model following the same procedures and form a new set of 10 portfolios. We set  $B = 12$ —a short window—to more closely follow the market dynamics, and  $F = 1$  to rebalance the portfolio at the end of each month. The top-decile portfolio based on this active ranking procedure using the forecast model leads to returns' superior performance, which will be further demonstrated in the empirical results from Section 4.

**Estimation procedure**

This section describes how the coefficients of the forecast model  $g(\cdot)$  or, equivalently  $\beta^t = (\beta_0^t, \beta_1^t, \beta_2^t, \dots, \beta_p^t)$  in Equation (4), are estimated. As  $\beta^t$  is estimated at each time  $t$ , we suppress the superscript  $t$  for  $\beta^t$  in the following discussion for simplicity. The ordinary least squares (OLS) estimator in Equation (4) can be efficiently obtained by minimizing the residual sum of the squares criterion, as:

$$\hat{\beta}^{ols} = \arg \min_{\beta} \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2. \tag{5}$$

where  $n$  is the total number of stocks.

When the model in Equation (4) considers numerous features, a well-known multicollinearity phenomenon may occur due to the high correlations among these features. For example, the three- and six-month cumulative returns may be strongly correlated. When correlated features are used, a ridge regression (Hoerl and Kennard 1970) is a popular approach to alleviate the multicollinearity problem. Specifically, the ridge estimator adds an  $L_2$  penalty term to the least square criterion as:

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right], \quad (6)$$

where  $\lambda$  is a tuning parameter that controls the magnitude of the penalty. By utilizing the bias-variance trade-off in the mean squared error (MSE), the ridge estimator reduces the MSE of the forecast from the OLS estimator when correlated features are used.

However, the ridge regression shrinks all the components of  $\beta$  proportionally compared with  $\hat{\beta}^{ols}$ , the estimated slope  $\hat{\beta}_j$  s at which all the features are non-zero. It is unclear whether the features with very small absolute  $\hat{\beta}_j$  s are less important in the model compared to those with larger absolute  $\hat{\beta}_j$  s. Thus, Tibshirani (1996) proposed a “lasso” estimator by replacing the  $L_2$  penalty in Equation (6) with the  $L_1$  penalty. Consequently, some  $\hat{\beta}_j$  s are shrunk to zero, which makes the lasso more than a tool for handling multicollinearity, it also becomes a tool for simultaneous feature selection. The lasso estimator takes the form of:

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right]. \quad (7)$$

To further improve lasso for high-dimensional problems in which the number of features is much larger than the number of observations (the number of stocks at time  $t$  in our discussion), Zou and Hastie (2005) combined the lasso and ridge penalty and proposed the elastic net (ENET) estimator as:

$$\hat{\beta}^{enet} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right]. \quad (8)$$

All three penalized estimators in Equations (6, 7 and 8) for the model in Equation (4) are commonly used in statistics and have also recently found adherence in finance and economics (Welsch and Zhou 2007; Bai and Ng 2008; Wang and Zhu 2010). Further, Hastie et al. (2009) offer a more in-depth review of these methods.

The lasso estimator is not very selective given a set of strong but correlated features and the ridge estimator is inclined to shrink the coefficients of correlated features toward each other. The compromise in the ENET estimator could allow highly correlated features to be averaged while encouraging a parsimonious model. Therefore, our empirical study proceeds with the elastic net estimator from Equation (8). Recent studies indicate the elastic net’s general usefulness in portfolio management (Shen et al. 2014; Montanari and Nguyen 2017).

The rest of this section will expatiate on how to obtain an effective solution  $\hat{\beta}^{enet}$  for Equation (8). Let  $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$ ; then solving  $\hat{\beta}^{enet}$  in equation (8) is equivalent to the optimization problem

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2$$

subject to

$$(1-\alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \leq t \text{ for some } t.$$

Therefore, the ENET estimator can be equivalently converted into the following form:

$$\beta^{enet} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left( \tilde{R}_{t+1,t+F}^i - \beta_0 - \sum_{j=1}^p \beta_j X_{j,t}^i \right)^2 + \lambda \left( (1-\alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \right) \right] \quad (9)$$

The ENET penalty is controlled by a constant  $0 < \alpha < 1$ , and bridges the gap between the lasso penalty ( $\alpha = 0$ ) and ridge penalty ( $\alpha = 1$ ). The tuning parameter  $\lambda$  controls the penalty's overall strength. By making  $\lambda$  sufficiently large, some of the coefficients are shrunk to exactly zero and the features are excluded in the forecast model.

Zou and Hastie (2005) show that given the penalty parameters  $\lambda$  one can solve Equation (9) based on an equivalent lasso-type problem based on a data augmentation process. Based on the LARS algorithm proposed by Efron et al. (2004), an efficient algorithm called LARS-EN can be employed to solve the entire elastic net solution path efficiently. Efron et al. (2004) proved that starting from zero, the lasso solution paths grow piece-wise linearly in a predictable way.

Hence, the computational efforts of obtaining an elastic net solution are equivalent to a single OLS fit. We refer the readers to Zou and Hastie (2005) and Efron et al. (2004) for more details. In our empirical study, this estimation is implemented using the "glmnet" package in the R software which effectively implements the above estimation procedure.

The tuning parameter  $\lambda$  is selected using a 10-fold cross-validation procedure. Specifically,  $\lambda$  is adjusted in small increments within a reasonable range to minimize the cross-validated forecast mean squared error (FMSE). To implement, the data is randomly split into 10 equal sized subsamples. Each time, a single subsample set is retained as validation data and the remaining nine subsamples are used as training data. For a given  $\lambda$ ,  $\hat{\beta}^{enet}$  is estimated using the training data and the FMSE is computed on the validation data using  $\hat{\beta}^{enet}$ . This process is repeated 10 times by holding out different subsamples as the validation data and the FMSE is computed each time. The cross-validated FMSE for each  $\lambda$  is the average of the 10 FMSEs. Ultimately, the optimal  $\lambda$  has the smallest cross-validated FMSE.

### Data and features

Although the Shanghai Composite Index was established in 1990, many of its security laws and regulations were introduced during the late 1990s, and the addition of new firms to the exchange significantly slowed after 2000. Therefore, our



empirical study focuses on the period since 2002. Our data includes the monthly returns for individual stocks and the Shanghai Composite Index returns from January 2002 to December 2016. Our work differs from that of Allen et al. (2017), which only considers Chinese A-share stocks, as we also include Chinese B-share stocks in our analysis for the following three reasons: First, the Shanghai Composite Index is a capitalization-weighted index, which tracks the daily price performance of all A- and B-shares listed on the Shanghai Stock Exchange. As will be observed, we use the index in constructing our features, and thus, it would be prudent for our study to include both A- and B-share stocks. Second, using both A- and B-share stocks increases the sample size  $n$  in the estimation procedure from Equation (9), which decreases the model forecast error in Equation (4). Third, it further validates our modeling framework's robustness to examine whether excluding B-share stocks in the portfolio would fundamentally change the strategy's behavior.

Our features  $\Omega_t^i = (X_{1,t}^i, X_{2,t}^i, \dots, X_{p,t}^i)$  used in Equation (4) are solely derived from stock prices without relying on any fundamental or accounting information. This set of features employs a variety of measures, including the cumulative monthly returns, Sharpe ratio, and the returns' skewness and kurtosis. Cross-sectional momentum has been well-documented in the US equity market; research has indicated that the stock price relates to its past price, and selecting stocks based on their past returns is intuitive. Jegadeesh and Titman (1993) found that buying well-performing stocks and selling short poor-performing stocks generates significant positive returns over 3- to 12-month holding periods. We follow these ideas by including the cumulative returns for twelve possible momentum evaluation look-back periods. As previously mentioned, the cumulative return is calculated by  $\tilde{R}_{t+1,t+F}^i = \prod_{j=1}^F (1 + R_{t+j}^i) - 1$ , where  $R_t^i$  is the monthly return of stock  $i$  at time  $t$ .

As the low volatility anomaly is also well-documented, we have also included volatility as a potential factor in our analysis. Blitz and van Vliet (2007) found that stocks with low historical volatility have high risk-adjusted returns in global markets, and the volatility effect cannot be explained by other factors, such as value, size, and momentum. Thus, we increase the model's accuracy by dividing this feature into two parts to reflect long-term and short-term volatility, respectively. Specifically, long-term volatility is the standard deviation calculated from the past 12 months of returns, while short-term volatility is calculated from the daily returns from the past 20 days. The market beta and alpha originated as concepts in the capital asset pricing model (Sharpe 1964; Lintner 1965), a cornerstone of finance. Frazzini and Pedersen (2014) proposed a strategy of buying low-beta stocks and selling short high-beta stocks to achieve a zero beta by adjusting positions. They demonstrate that this strategy produces significant and positive risk-adjusted returns. Additionally, we employ the current drawdown of a stock, measured as the percentage difference between the stock's most recent price and its 52-week high. Chen and Yu (2016) discovered that this measure has certain exclusive unpriced information in the cross-sectional pricing of stocks.

To summarize, we grouped our features into two categories: return-based and statistics-based. Table 1 lists all the features used in our analysis. All features are standardized before being their fit into the statistical learning model.

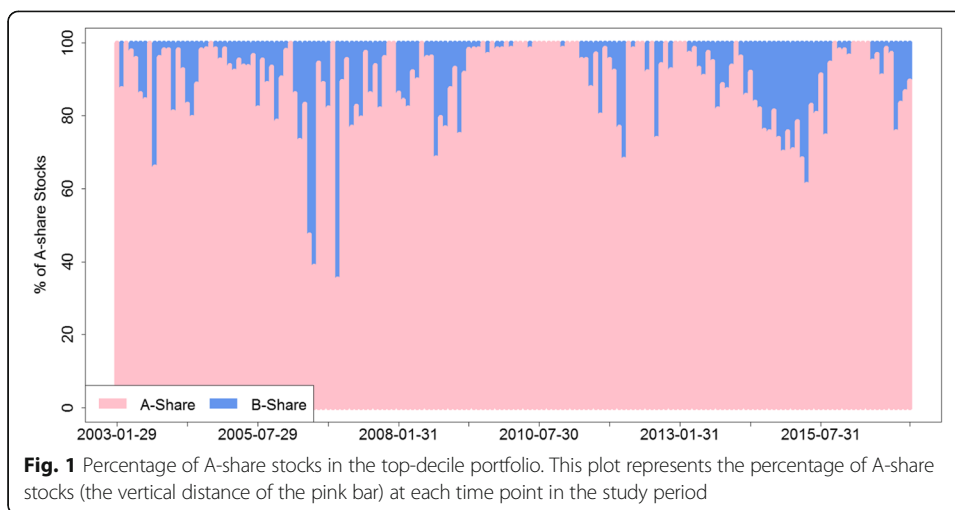
**Table 1** Features used in the study

Features	Symbol
Return based:	
1. One-month return	R1
2. Two-month cumulative return	R2
3. Three-month cumulative return	R3
4. Four-month cumulative return	R4
5. Five-month cumulative return	R5
6. Six-month cumulative return	R6
7. Seven-month cumulative return	R7
8. Eight-month cumulative return	R8
9. Nine-month cumulative return	R9
10. Ten-month cumulative return	R10
11. Eleven-month cumulative return	R11
12. Twelve-month cumulative return	R12
Statistics based:	
13. Standard deviation calculated from the past 12 months of returns	SD
14. Skewness calculated from the past 12 months of returns	Skew
15. Kurtosis calculated from the past 12 months of returns	Kurt
16. Sharpe ratio calculated from the past 12 months of returns	Sharpe
17. Alpha with respect to the market calculated from the past 12 months of returns	Alpha
18. Market beta calculated from the past 12 months of returns	Beta
19. Correlation with index calculated from the past 12 months of returns	Corr
20. Short-term volatility calculated from the daily returns in the past 20 days	Vol
21. Distance/drawdown from the high watermark in the past 12 months	DD

### Empirical results

We follow the steps described in Section 2.2, and set the look-back rolling window  $B$  as 12 months, such that the actual performance evaluation period is from January 2003 to December 2016. In setting the look-forward window  $F$  to one month, we evaluate three trading strategies based on the forecasted one-month return from the elastic net model. First, we apply our strategy to buy the top-decile stocks while selling short the bottom-decile stocks. We do this because asset pricing literature often uses a long-short dollar-neutral strategy (Jegadeesh and Titman 1993; Frazzini and Pedersen 2014) to test market efficiency and examine certain features' usefulness. We denote the portfolio constructed using this strategy as the Equity-Market-Neutral (EMN-ENET) portfolio, as it seeks to exploit investment opportunities unique to some specific group of stocks while maintaining a neutral exposure to broad groups of stocks. We then evaluate our strategy's performance by only buying the top-decile stocks that adapt to the regulations on short-sales in the Chinese stock market. We denote this portfolio as the Long-ENET portfolio.

Domestic investors are only allowed to invest in A-share stocks. As Fig. 1 illustrates, the percentage of A-share stocks in the top-decile portfolio selected by the proposed forecast approach may contain up to 60% of B-share stocks. Therefore, it is beneficial to examine whether the A-share stocks the forecast model selects can independently generate good performance. We call the portfolio constructed only with selected



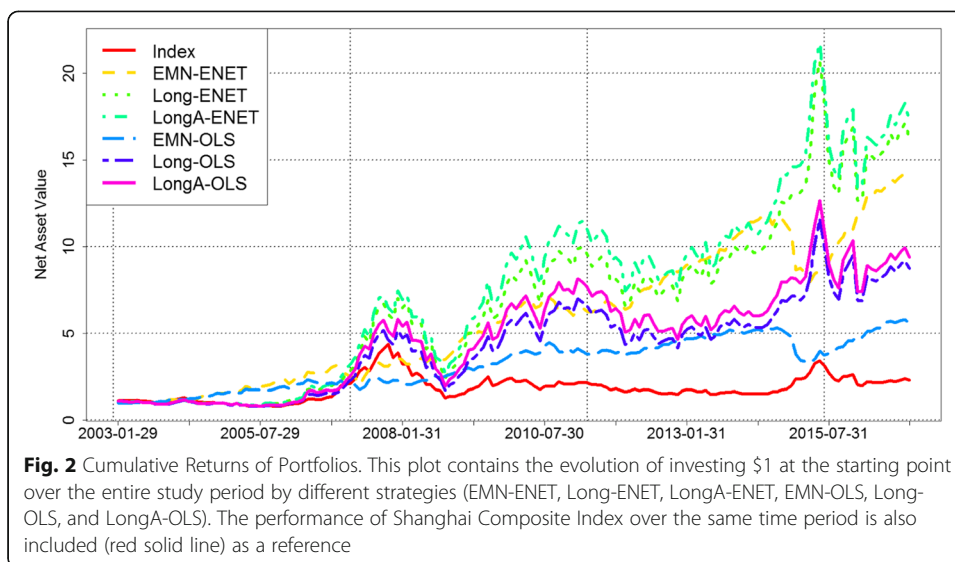
A-share stocks the LongA-ENET portfolio. To further demonstrate the advantage of using the ENET estimator over the OLS estimator, we construct three parallel portfolios using the OLS estimator instead of the ENET estimator when estimating Equation (4) and denote these as the EMNOLS, Long-OLS, and LongA-OLS portfolios, respectively.

### Performance assessment

We then assess the proposed investment mechanism's performance by comparing the cumulative returns over the evaluation period through an investment of one dollar (or RMB) in January 2003 to the corresponding cumulative returns from the Shanghai Composite Index during the same period.

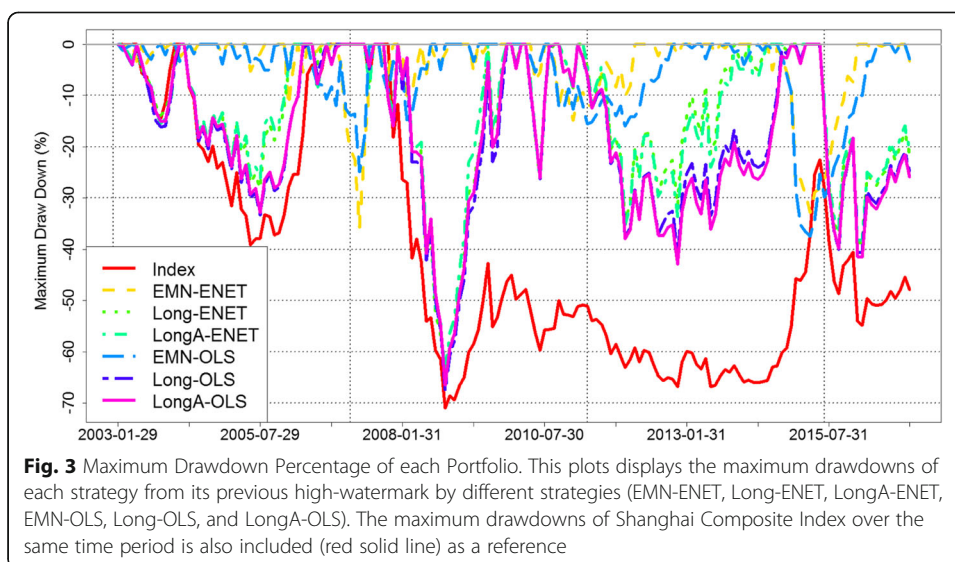
Figure 2 clearly indicates that despite the underlying index's lackluster performance, all three ENET strategies based on the proposed forecast models generate superior overall performance. The two long-only portfolios rose and fell with the index during the major 2005–2008 and 2014–2015 cycles. However, our strategies still identify well-performing stocks in the 2010–2013 timespan, which achieved new highs while the index failed to post any meaningful gains. The EMN-ENET portfolio has the smoothest overall cumulative return curve, and especially during the 2008 global financial crisis, because the proposed forecast model effectively detects changes in the market environment and successfully selects underperforming stocks to benefit from selling short. Additionally, we observe that the stock selection strategies still significantly outperform the index where the OLS estimator is used instead of the ENET estimator, but are not as effective as with the ENET approach. This demonstrates the value of considering a more sophisticated ENET estimator over a simple OLS estimator.

We delineate each strategy's downside risk by comparing the drawdowns of each portfolio from their high watermarks. At each time  $t$ , each strategy's high watermark  $M_t$  is the highest value of historical cumulative returns, and can be calculated as  $M_t = \max_{k=1, \dots, t} \{\tilde{R}_{1,k}\}$  where  $\tilde{R}_{1,k} = \prod_{j=1}^k (1 + R_j) - 1$  and  $R_j$  is the portfolio's return at time  $j$ . The portfolio's drawdown at time  $t$  is defined as  $(\tilde{R}_{1,t} - M_t) / M_t$ . Figure 3 compares the percentage of drawdowns for different strategies. In the plot, the point in time



at which the time series lines last touch the 0-reference line (the gray-dashed line at the top) is the time when each portfolio reaches its historical highest cumulative return. Given its hedging ability, the EMN-ENET portfolio has the lowest overall drawdown percentage and successfully avoided the market turmoil in 2008. Although the Long-ENET and LongA-ENET strategies suffered in 2008 and in the overall bearish Chinese stock market in 2015, they recovered much faster than the index to achieve a new historical high watermark. This is because the proposed forecast model quickly updates the stock rankings in different market conditions and adjusts the allocations in the portfolio accordingly. The relative comparisons between pairs of the EMN, Long, and LongA strategies hold the same for using the OLS estimator instead of the ENET estimator. However, if the OLS estimator is used instead of the ENET estimator, the drawdowns are consistently more severe.

For further insights, Table 2 reports the annualized return (AR); the standard deviation (Std); the maximum monthly return during the evaluation period (Best MR); the



**Table 2** Performance measures during 2003–2015

Statistics	Index	EMN-ENET	Long-ENET	LongA-ENET	EMN-OLS	Long-OLS	LongA-OLS
AR (%)	6.08	20.59	22.03	22.52	13.12	16.75	17.34
Std (%)	28.88	19.08	35.27	35.58	16.80	34.81	35.40
Best MR (%)	27.45	25.16	35.24	41.14	20.92	29.38	38.31
Worst MR (%)	−24.63	−20.38	−26.46	−26.46	−21.20	−27.18	−28.35
ShR	0.21	1.08	0.67	0.63	0.78	0.48	0.49
MD (%)	70.97	35.75	66.58	64.74	37.41	67.39	66.46
CarR	0.09	0.58	0.33	0.34	0.35	0.25	0.26
Skewness	−0.21	−0.06	0.15	0.20	−0.17	−0.01	0.09
Kurtosis	1.10	3.39	1.08	1.34	3.77	0.59	1.01
Correlation	1.00	−0.10	0.82	0.83	−0.12	0.84	0.83

minimum monthly return during the evaluation period (Worst MR); the Sharpe ratio (ShR); the maximum drawdown (MD); the Calmar ratio (CalR), which is calculated as the ratio of annualized returns over the maximum drawdown; skewness (Skew); kurtosis (Kurt); and the correlation between each portfolio and the Shanghai Composite Index (Correlation). According to Table 2, the EMN-ENET, Long-ENET, and LongA-ENET strategies achieved 20.59%, 22.03%, and 22.52% annualized returns during the evaluation period, respectively, while the index only registered an annualized return of approximately 6%. More importantly, all three strategies delivered higher risk-adjusted measurements as indicated by the Sharpe ratio and the Calmar ratio. The short look-forward window ( $F = 1$ ) is proven as highly effective, and especially in dealing with the 2008 financial crisis, during which time the EMN-ENET strategy generated a positive 13.8% return in 2008, while the index itself lost more than half its value in the same year. The success in identifying winning and losing stocks also generates a portfolio that negatively correlates with the Shanghai Composite Index. In contrast to the EMN-ENET strategy, the two long-only strategies are less stable with slightly larger annualized standard deviations than the Shanghai Composite Index. However, they still deliver higher excess returns than the index. Table 3 presents the annual returns of each strategy from 2003 to 2015. In 2007, the highest annual return among the proposed strategies was 216%, which is nearly twice as large as the index's highest annual return during the entire evaluation period. Hence, the proposed strategies achieve much higher Sharpe and Calmar ratios than the index, with similar risks but much greater upside potential.

We also observe whether B-share stocks are included in the top-performing portfolios makes a small difference. The previously mentioned results indicate that the two Long and LongA strategies exhibit virtually the same performance in terms of their risk-adjusted returns. This is advantageous to investors in the Chinese stock market who are regulated regarding short-sales, and who are only allowed to invest in A-share stocks. One way to interpret the similarity between the two long-only strategies is to handle the A-share-only restriction as a “pseudo-random” sampling process from all stock pools. The A- and B-share stocks selected by the forecast models bear similar characteristics, and random samples obtained from the top decile portfolio would generate similar good performance as long as the sample size is not too small.

**Table 3** Annual Returns for 2003–2015

Year	Index	EMN-ENET	Long-ENET	LongA-ENET	EMN-OLS	Long-OLS	LongA-OLS
2003	-0.19	17.47	-9.95	-11.40	19.50	-11.61	-10.29
2004	-15.40	35.52	-3.63	-2.91	34.94	-3.62	-2.54
2005	-8.33	48.73	8.62	9.60	22.32	-5.01	-5.11
2006	130.43	26.92	117.34	142.38	15.73	93.60	109.85
2007	96.66	18.04	216.39	201.58	0.68	205.97	209.02
2008	-65.39	13.81	-56.19	-53.52	19.15	-57.98	-56.57
2009	79.98	60.94	172.12	177.43	40.01	165.45	166.35
2010	-14.31	-2.83	10.52	10.42	7.83	12.35	12.52
2011	-21.68	0.53	-29.17	-30.12	-7.91	-32.86	-33.09
2012	3.17	36.69	23.79	21.43	22.00	12.83	8.84
2013	-6.75	23.70	19.94	21.00	12.14	11.88	12.91
2014	52.87	-19.72	33.62	33.69	-26.26	29.72	31.02
2015	9.41	27.18	29.86	22.85	21.17	32.95	27.07
2016	-12.31	25.04	-3.84	-4.16	22.07	-7.81	-9.33

Tables 2 and 3 reveal that the OLS strategies generally have smaller volatility, but with much smaller annualized returns. Additionally, the best monthly return is inferior to that with the ENET strategies, and the worst month return is respectively worse. Consequently, they deliver lower Sharpe and Calmar ratios. However, regardless of whether the ENET or OLS methods are used, the statistical modeling approach consistently outperforms the index. As the ENET strategies are proven superior to the OLS strategies, our following analysis will focus on the former.

### Robustness analysis

We assess the proposed ENET strategies' consistency and robustness by recalculating the previously mentioned performance measures over a longer time period, of five-year rolling windows instead of one-year windows. Specifically, we compute the performance measures reported in Table 2 for five overlapping periods: 2003–2007, 2005–2009, 2007–2011, 2009–2013, and 2011–2015.<sup>2</sup>

Table 4 reports the average of the performance measures over the five five-year overlapping rolling time windows. With a more diversified risk premium over a longer evaluation window, all the performance measures are improved, especially for the proposed Long-Only and Long-A-Only strategies in which the average annualized returns are significantly higher over a five-year period with risks similar to that for the one-year period. As a result, their correlations between the index remain nearly unchanged while the Sharpe ratios are much higher and the maximum drawdowns are lower.

As a comparison to Table 3, we report the average annualized returns over the five five-year rolling windows in Table 5. Averaging over a longer period reveals the proposed strategies' apparent stability and advantages. When experiencing an overall bullish market, as in 2003–2007, the EMN strategy generated return performance similar to the index but with a higher Sharpe ratio, while the two long-only strategies generated much higher returns compared with the index.

**Table 4** Average performance measures during 2003–2015 for 5-year rolling windows

Statistics	Index	EMN-ENET	Long-ENET	LongA-ENET
AR (%)	11.20	22.15	33.08	34.07
Std (%)	29.56	20.36	36.98	37.32
Best MR (%)	22.28	19.46	29.97	31.48
Worst MR (%)	−20.72	−16.04	−19.47	−19.83
ShR	0.38	1.10	0.89	0.89
MD (%)	51.37	28.60	46.51	45.37
CalR	0.25	0.84	0.80	0.85
Skewness	−0.08	−0.25	0.26	0.23
Kurtosis	0.69	2.85	0.46	0.62
Correlation	1.00	−0.07	0.81	0.82

During 2007–2011 when a severe market drawdown occurred, all three of the proposed strategies still generated positive excess returns while the market index experienced a loss with a negative Sharpe ratio.<sup>3</sup> As the market slowly recovered during 2009–2013 the three proposed strategies recovered much faster, and all delivered over 20% annualized returns during that five-year window. In conclusion, the long-term performance analysis results reinforce the observation that the proposed forecast-based strategies not only consistently deliver significant excess returns over the market index, but also self-adapt during both bullish and bearish Chinese stock markets.

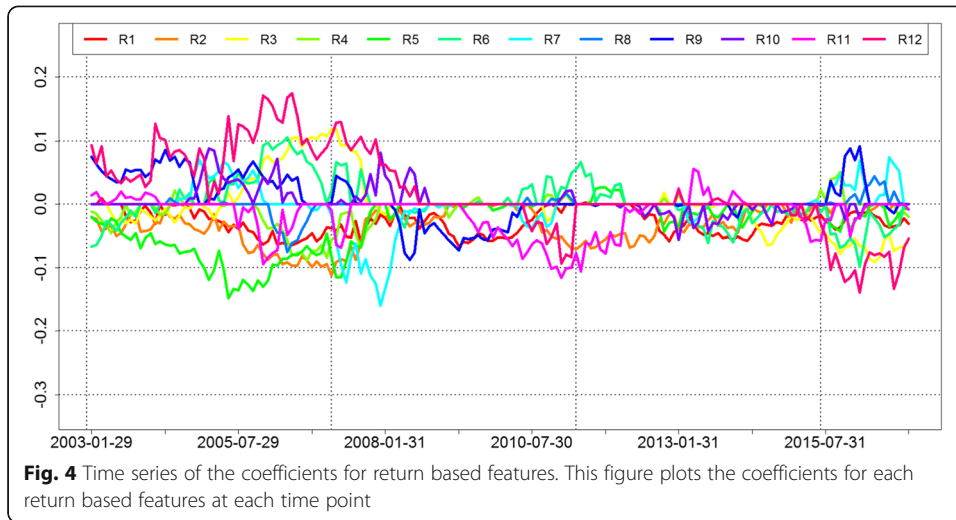
#### Feature analysis

Given the proposed strategies' superior portfolio performance, it is informative to examine how each feature is weighted in forecasting stock returns. Figures 4 and 5 illustrate the behavior of the estimated coefficients  $\beta_j^t$  corresponding to both return- and statistics-based features (defined in Table 1) in Equation (4) at each time  $t$  from January 2003 to December 2015. These coefficient paths illustrate the corresponding shifts of market regimes as their signs changed in the sample period.

Some compelling behaviors can be observed from the coefficient evolution paths for the different return-based features. The coefficients of the medium to long term momentum features R3, R6 and R12 were positive in the time period before 2009 which is in agreement with the cross-sectional momentum anomaly discussed in Jegadeesh and Titman (1993). However, their signs flipped in the following years signaling a potential market structure change which correctly reflect the behavior of the Chinese stock market. Meanwhile, the coefficient of the one-month momentum R1 remains to be negative in the entire time period. This phenomenon is in

**Table 5** Average annualized return over a 5-year rolling window

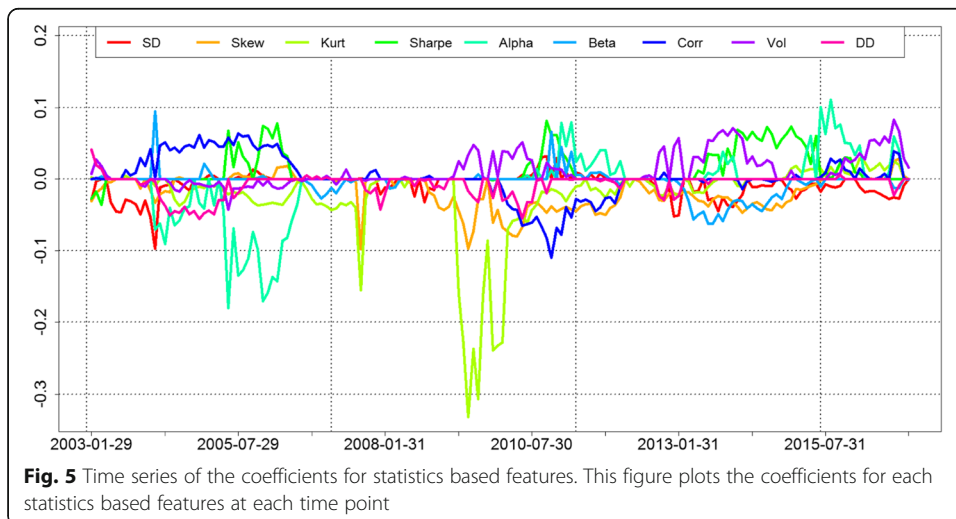
Period	Index	EMN-ENET	Long-ENET	LongA-ENET
2003–2007	31.12	28.90	47.67	49.49
2005–2009	20.94	32.48	54.85	59.52
2007–2011	−3.84	16.13	24.18	24.58
2009–2013	3.05	21.60	25.90	25.76
2011–2015	4.74	11.66	12.79	11.01



agreement with a short-term mean reversion relationship of stock returns described by Jegadeesh and Titman (1990).

The individual statistics based features reveal the characteristics of the stocks favored to generate high excess (above market) returns. It is preferable to select stocks with higher past Sharpe ratios, smaller tails as reflected by the kurtosis, lower volatility, and more negative skewness. The first three features are preferred because they generally indicate lower risk while the preference for negative skewness may indicate a risk premium compensating for risk that is systematically biased downwards. It is also preferable to select stocks not close to their 12-month high watermark. The coefficient paths for the statistics based features support these preferences. For example, the coefficients of Skew and DD are almost negative over the entire evaluation period; and the coefficient on the kurtosis effectively reflects structural changes in the Chinese stock market.

As described in Section 2.3, the elastic net estimator from Equation (8) implements feature selection. This advantage enables a convenient examination of how frequently





each feature contributes to the final portfolio's composition. We achieve this goal by constructing a Feature Importance (FI) score as a simple statistic to measure each feature's degree of importance. If a feature  $X_{j,t}$  in Equation (4) does not sufficiently contribute to the forecast model at time  $t$ , its corresponding coefficient  $\beta_j^t$  is shrunk to zero by the elastic net estimator and is excluded from the active feature set at time  $t$ . Hence, we define the FI score for a feature  $X_j$  as  $FI_j = N/T$  where  $N$  is the number of occurrences of feature  $X_j$  during a specific time window and  $T$  represents the number of total time periods within a given time span. If the FI score of a feature  $X_j$  equals zero, this indicates that feature  $X_j$  is never "useful" in selecting stocks over  $T$  periods of time. In contrast, if the FI score of a feature  $X_j$  equals one, that indicates that the feature is vital in building the portfolio. Therefore, the larger the FI score, the more important (useful) the feature.

The evaluation period from 2003 to 2016 is divided into seven equally spaced periods as Table 6 indicates. Hence,  $T$  equals 24 in each time period. The average FI score during the time from 2007 to 2012 is lower than other time periods meaning that fewer features are useful during this period when the economy entered a global financial crisis and subsequently recovered. Based on the overall importance in Table 6 (the last column), the one-month (R1) and two-month (R2) cumulative returns and Kurtosis (Kurt) are the three most important features over time while the eight-month cumulative return showed up the least frequently.

**Table 6** Feature importance scores in different time periods

Factors	2003–2004	2005–2006	2007–2008	2009–2010	2011–2012	2013–2014	2015–2016	2003–2016
R1	0.71	1.00	1.00	0.96	0.50	1.00	1.00	0.88
R2	1.00	1.00	1.00	0.46	1.00	0.88	0.46	0.83
R3	0.75	0.96	0.50	0.17	0.08	0.58	1.00	0.58
R4	0.67	0.63	0.46	0.29	0.17	0.04	0.38	0.38
R5	1.00	1.00	0.50	0.21	0.46	0.67	0.67	0.64
R6	0.58	1.00	0.54	0.33	0.54	0.92	0.83	0.68
R7	0.25	0.63	0.92	0.42	0.00	0.21	0.67	0.44
R8	0.08	0.42	0.00	0.13	0.13	0.00	0.42	0.17
R9	0.92	0.88	0.67	0.67	0.04	0.13	0.42	0.53
R10	0.17	0.75	0.54	0.00	0.46	0.71	0.17	0.40
R11	0.54	0.54	0.17	0.83	0.75	0.33	0.63	0.54
R12	1.00	1.00	0.79	0.29	0.13	0.21	0.92	0.62
SD	0.79	0.46	0.42	0.33	0.54	0.79	0.83	0.60
Skew	0.50	0.50	0.38	1.00	0.92	1.00	0.50	0.68
Kurt	0.54	1.00	0.50	0.88	0.79	0.54	0.96	0.74
Sharpe	0.13	0.58	0.00	0.42	0.13	0.83	0.50	0.37
Alpha	0.42	0.83	0.04	0.17	0.54	0.25	0.58	0.40
Beta	0.13	0.33	0.38	0.13	0.50	0.96	0.33	0.39
Corr	0.75	0.79	0.21	0.58	0.58	0.33	0.67	0.56
Vol	0.58	0.79	0.25	0.83	0.42	0.79	0.79	0.64
DD	0.71	0.38	0.46	0.58	0.42	0.08	0.13	0.39
Average FI	0.58	0.74	0.46	0.46	0.43	0.54	0.61	0.55

## Conclusions

This study constructed a cross-sectional statistical forecast model for stock selection in the Chinese stock market. Based on the forecast of future cumulative returns, the proposed approach allows investors to identify stocks that are likely to perform well, and to construct corresponding portfolios. Although the Chinese stock market collectively cannot generate satisfactory results, our empirical results indicate that it is still possible to generate significant excess returns through an active, quantitative stock selection process. Additionally, and regarding the features used in this study, far fewer useful features were observed during the 2008 financial crisis than in other periods of time. Meanwhile, the frequent occurrence and negative signs in the one-month return (R1) feature strongly indicate a short-term mean-reversion in the Chinese stock market. Further, despite the fact that the forecast models are constructed using both A- and B-share stocks, solely investing in A-share stocks from the top-decile portfolio can still yield incredibly good performance.

## Endnotes

<sup>1</sup>We choose to rebalance at the end of the look-forward window following Moskowitz et al. (2010), as the forecast model is built to predict cumulative returns in the same time window.

<sup>2</sup>We chose to rollover with a 2-year increment between each 5-year evaluation window because this allows us to evaluate long-term performance over equally spaced time windows for the entire 14-year period.

<sup>3</sup>We obtained complete performance measures for each five-year period; although these results are not included in this paper due to space limitations, they are available upon request.

## Acknowledgments

The authors thank the editor, the associate editor, and the two anonymous referees for constructive remarks which greatly improved the content and exposition of this paper. Michael L. Tindall-The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

## Funding

This study is currently not supported by any other funding agency and we do not have any financial interests that may influence the research.

## Availability of data and materials

The data used for our analysis are provided by the Wind Database5, a leading financial database and software services provider that provides financial securities data, covering stocks, funds, bonds, foreign exchange, insurance, futures, financial derivatives, spot trade, macroeconomics, financial news and other fields.

## Authors' contributions

All the listed authors have made various contributions to the development of this paper. The first and second author conducted all the empirical analysis and made significant contribution to the composition of the paper. The third and fourth author contribute significantly by providing and preparing the data for the study. They also provided constructive discussions to improve the exposition of the paper. The last author contributed by providing some very meaningful suggestion that helped improve the overall quality of the paper. The manuscript is not under simultaneous review, and no other closely related manuscripts exist. We do not have any financial interests that may influence the research. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

## Publisher's Note

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

**Author details**

<sup>1</sup>Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX, USA.

<sup>2</sup>Federal Reserve Bank of Dallas, Financial Industry Studies Department, Dallas, TX, USA. <sup>3</sup>School of Business Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

Received: 19 September 2018 Accepted: 15 April 2019

Published online: 29 April 2019

**References**

- Allen, F., Qian, J., Shan, S. C., and Zhu, J. L. (2017). Explaining the disconnection between china's economic growth and stock market performance. *Wharton and SAIF Working Paper*.
- Ang A, Hodrick RJ, Xing Y, Zhang X (2006) The cross-section of volatility and expected returns. *J Financ* 61:259–299.
- Bai J, Ng S (2008) Forecasting economic time series using targeted predictors. *J Econ* 146(2):304–317.
- Blitz D, van Vliet P (2007) The volatility effect: lower risk without lower return. *J Portf Manag* 34(1):102–113.
- Carhart MM (1997) On persistence in mutual fund performance. *J Financ* 7(1):57–82.
- Chen L-W, Yu H-Y (2016) Nearness to the 52-Week High and Low Prices, Past Returns, and Average Stock Returns (June 17, 2016). 29th Australasian Finance and Banking Conference 2016. Available at SSRN: <https://ssrn.com/abstract=2797149> or <http://dx.doi.org/10.2139/ssrn.2797149>.
- Efron B, Hastie T, Johnstone I (2004) Least angle regression. *Ann Stat* 10:407–499.
- Fama EF, French KR (1992) The cross-section of expected stock returns. *J Financ* 47(2):427–465.
- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116:1–22.
- Fama EF, MacBeth JD (1973) Commonality in the determinants of expected stock returns. *J Financ Econ* 81(3):607–636.
- Frazzini A, Pedersen LH (2014) Betting against beta. *J Financ Econ*:171–125.
- Gu, S., Kelly, B., and Xiu, D. (2019). Empirical asset pricing via machine learning. *working paper*.
- Harvey CR, Liu Y, Zhu H (2016) and the cross-section of expected returns. *Rev Financ Stud* 29(1):5–68.
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, 2nd edn. Springer, New York.
- Haugen RA, Baker NL (1996) Commonality in the determinants of expected stock returns. *J Financ Econ* 41(3):401–439.
- Hoerl AE, Kennard RW (1970) Ridge regression: applications to nonorthogonal problems. *Technometrics* 12(1):69–82.
- Hou K, Xue C, Zhang L (2015). Digesting Anomalies: An Investment Approach. *Rev Financ Stud* 28(3):650–705
- Jegadeesh N, Titman S (1990) Evidence of predictable behavior of security returns. *J Financ* 45(3):881–898.
- Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. *J Financ* 48(1):65–91.
- Kang J, Liu M-H, Ni SX (2002) Contrarian and momentum strategies in the China stock market: 1993–2000. *Pac Basin Financ J* 10(3):243–265.
- Kling G, Gao L (2008) Chinese institutional investors' sentiment. *J Int Financ Mark Inst Money* 18(4):374–387.
- Li B, Hoi SCH (2015) Online portfolio selection: principles and algorithms. Crc Press.
- Li B, Zhang D, Zhou Y (2017) Do trend following strategies work in chinese futures markets? *J Futur Mark* 37(12):1226–1254.
- Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev Econ Stat* 47(1):13–37.
- Mohanram PS (2005) Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Rev Acc Stud* 10:133–170.
- Montanari A, Nguyen P-M (2017) Universality of the elastic net error. In: *Information theory (ISIT)*. 2017 IEEE International Symposium, pp 2338–2342 IEEE.
- Moskowitz TJ, Ooi YH, Pedersen LH (2010) Time series momentum. *J Financ Econ* 104:228–250.
- Rapach D, Zhou G (2013) Forecasting stock returns. *Handbook of economic forecasting*:39328–39383.
- Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *J Financ* 19(3):425–442.
- Shen W, Wang J (2017) Portfolio selection via subset resampling. *AAAI*, pp 1517–1523.
- Shen W, Wang J, Ma S (2014) Doubly regularized portfolio with risk minimization. *AAAI*, pp 1286–1292.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 58(1):267–288.
- Varian HR (2014) Big data: new tricks for econometrics. *J Econ Perspect* 28(2):3–28.
- Wang F, Xu Y (2004) What determines chinese stock returns? *Financ Anal J* 60(6):65–77.
- Wang L, Zhu J (2010) Financial market forecasting using a two-step kernel learning method for the support vector regression. *Ann Oper Res* 174(1):103–120.
- Welsch RE, Zhou X (2007) Application of robust statistics to asset allocation models. *Statistical J* 5(1):97–114.
- Wu W, Chen J, Yang Z(B), Tindall ML (2008) A Cross-Sectional Machine Learning Approach for Hedge Fund Return Prediction and Fund Selection (August 16, 2018). Available at SSRN: <https://ssrn.com/abstract=3238466> or <http://dx.doi.org/10.2139/ssrn.3238466>.
- Yuan R, Xiao JZ, Zou H (2008) Mutual funds' ownership and firm performance: evidence from China. *J Bank Financ* 32(8): 1552–1565.
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B* 67(2):301–320.