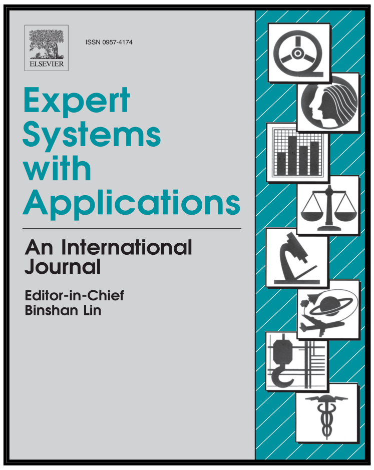
Journal Pre-proof



Effect of dimensionality reduction on stock selection with cluster analysis in different market situations

Jingti Han, Zhipeng Ge

PII: S0957-4174(20)30052-X
DOI: https://doi.org/10.1016/j.eswa.2020.113226
Reference: ESWA 113226

To appear in: *Expert Systems With Applications*

Received date: 14 February 2019
Revised date: 16 January 2020
Accepted date: 19 January 2020

Please cite this article as: Jingti Han, Zhipeng Ge, Effect of dimensionality reduction on stock selection with cluster analysis in different market situations, *Expert Systems With Applications* (2020), doi: https://doi.org/10.1016/j.eswa.2020.113226

Highlights

1. Dimensionality reduction hardly improves the Sharpe ratio of stock selection in sideways
2. The advantage of dimensionality reduction is mainly reflected in trend situations
3. A stock-selection rotation strategy with and without dimensionality reduction is proposed
4. The Sharpe ratio of the rotation strategy is higher than that of benchmark strategies

Effect of dimensionality reduction on stock selection with cluster analysis in different market situations

Jingti Han[a,b], Zhipeng Ge[a,b,*]

[a] *School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, PR China*
[b] *Institute of Fintech, Shanghai University of Finance and Economics, Shanghai 200433, PR China*

Abstract

Dimensionality reduction is inevitable in stock selection with cluster analysis. Considering relations among dimensionality reduction, noise trading, and market situations, we empirically investigate the effect of dimensionality-reduction methods–principal component analysis, stacked autoencoder, and stacked restricted Boltzmann machine–on stock selection with cluster analysis in different market situations. Based on the index fluctuation, the market is divided into sideways and trend situations. For the CSI 100 and Nikkei 225 constituent stocks, experimental results show that: (1) in sideways situations, dimensionality reduction hardly improves the performance of stock selection with cluster analysis; (2) the advantage of dimensionality reduction is mainly reflected in trend situations, but whether it is in an up or down trend depends on the market analyzed. More importantly, according to the above findings and assuming that the dimensionality-reduction effect will continue, we propose a rotation strategy with and without dimensionality reduction. The results of experiments show that the proposed rotation strategy outperforms the stock market indices as well as the stock-selection strategies based on dimensionality reduction and cluster analysis. These findings offer practical insights into how dimensionality reduction can be efficiently used for stock selection.

Keywords: Stock selection, Dimensionality reduction, Market situation, Rotation strategy, Deep learning

1. Introduction

Stock selection is a crucial issue in investment management, which determines the return of stock investments (Markowitz, 1952; Ren et al., 2017). There are various stock-selection strategies, including multi-factor models (Carvalho

*Corresponding author
Email addresses: `hanjt@mail.shufe.edu.cn` (Jingti Han), `lzgezhipeng@126.com` (Zhipeng Ge)

et al., 2010; Fama and French, 2018), momentum and contrarian strategies (Grinblatt et al., 1995; Cooper et al., 2004), style rotation strategies (Lucas et al., 2002; Ahmed et al., 2002), volatility strategies (Chong and Phillips, 2012; Hsu and Li, 2013), and behavior biases strategies (Huang et al., 2011). Among these strategies, multi-factor models are the most studied, mainly including the Fama-French three-factor model (Fama and French, 1992), the Fama-French five-factor model (Fama and French, 2017), factor models based on investor attention (Li and Yu, 2012), and factor models based on fundamental and technical analysis (Peachavanish, 2016). Investors can use these models to analyze stock characteristics from different perspectives. If stock characteristics last for a period, investors would obtain a higher benefit from analyzing stock characteristics than from random selection.

Stock selection with cluster analysis has attracted investors and researchers' attention (Hu et al., 2018; Iorio et al., 2018). An analysis of stock clusters for the Thai stock market found a higher return on stock selection with cluster analysis than without it (Peachavanish, 2016). Investors can use cluster analysis based on various characteristics. Da Costa Jr et al. (2005) employed cluster analysis with fundamental and technical factors to classify stocks and analyzed the return-risk ratio for each cluster. Importantly, investors can detect the relation among stocks by cluster analysis. Brida and Risso (2010) analyzed the hierarchical structure of German stock markets. Tabak et al. (2010) explored topological properties of Brazil stock markets. Dose and Cincotti (2005) and Silva and Marques (2010) found that stock selection accounting for relations among stocks determined the excess return of enhanced index tracking portfolio. With relation findings, investors can select a variety of stocks. Nanda et al. (2010) used K-means, self organizing maps (SOM), and fuzzy C-means to select stocks, and then employed the Markowitz theory for stock allocation. They found that stock selection with cluster analysis can improve portfolio performance. Baser and Saini (2015) used K-means, K-medoids, and fast K-means to select stocks, and analyzed the efficient frontier for each cluster. From the above literature, it is concluded that stock selection with cluster analysis provides the following advantages: (1) investors or researchers can use many effective characteristics to analyze stocks and further construct portfolios; (2) combining stock characteristics, investors can well detect the relation among stocks; (3) investors can select diverse stocks from different clusters, which is beneficial to reduce the systemic risk of portfolios; (4) investors can calculate the allocation of selected stocks rather than of all stocks in the market quickly.

In practice, the curse of dimensionality is inevitable in cluster analysis with high-dimensional data (Ding et al., 2002; Verleysen and François, 2005; Tajunisha and Saravanan, 2010). Steinbach et al. (2004) discussed its challenges, and Parsons et al. (2004) reviewed cluster algorithms for high-dimensional data. Before stock selection with cluster analysis, Fulga et al. (2009) proposed principal component analysis (PCA) to reduce the effect of dimensions and found this strategy can produce useful results for portfolio optimization. Unfortunately, conventional methods, including principal component analysis (Jolliffe and Cadima, 2016), linear local embedding (Roweis and Saul, 2000), and Sammon mapping (Sammon, 1969), have obvious drawbacks of the assumptions of linear or local manifold relations. Dimensionality-reduction methods based on neural networks (Cai et al., 2012), such as the stacked autoencoder (SAE) (Hinton and Salakhutdinov, 2006) and stacked restricted Boltzmann machine (SRBM) (Hinton et al., 2006), have been widely used in image, speech, and finance (LeCun et al., 2015; Li et al., 2015; Heaton et al., 2017; Chong et al., 2017). These non-parametric methods can learn nonlinear relations and have strong self-learning and fault-tolerance ability. Therefore, it is significant and urgent to investigate the effect of non-parametric methods on stock selection with cluster analysis.

However, there are complex relations among dimensionality reduction, noise trading, market volatility, and market situations. According to Kirkpatrick and Dahlquist (2010), investors can judge market situations by volatility, which arises by the interaction of fundamental and noise trading (Verma and Verma, 2007). As we all know, dimensionality reduction is equivalent to signal compression, and can retain the main information while decreasing the noise in the data (Van Der Maaten et al., 2009), so our concern is about the effect of dimensionality reduction on stock selection with cluster analysis in different market situations. In addition, given this effect, can a significant investment be proposed to improve the performance of stock selection?

In this paper, we first divide market data into training and validation sets based on time. And then we train three dimensionality-reduction methods in the training set, including principal component analysis, stacked autoencoder, and stacked restricted Boltzmann machine. Further, we utilize trained dimensionality-reduction methods to reduce stock characteristics in the validation set, and evaluate the effect of trained dimensionality-reduction methods on stock selection with cluster analysis in different market situations which are divided based on the index fluctuation. The decision pipeline of this study is shown in Fig. 1. For the China Securities 100 Index (CSI 100) and Nikkei 225 constituent stocks, the results indicate that the advantage of dimensionality reduction is mainly reflected in trend situations, but whether it is in an up or down trend mainly depends on the market analyzed. Furthermore, based on these findings and assuming the effect of dimensionality reduction will continue, we propose a rotation strategy with and without dimensionality reduction. The findings of a series experiments show that the proposed rotation strategy outperforms the stock market indices, the stock selection with dimensionality reduction and cluster analysis, and stock selection with cluster analysis.

The rest of this paper is organized as follows. In section 2, we introduce three dimensionality-reduction methods and a stock-selection strategy with cluster analysis. In section 3, we analyze the effect of dimensionality reduction on stock selection in trend and sideways situations for the CSI 100 and Nikkei 225 constituent stocks. A stock-selection rotation strategy based on the effect of dimensionality reduction is proposed in section 4, and conclusions and discussions are summarized in the last section.

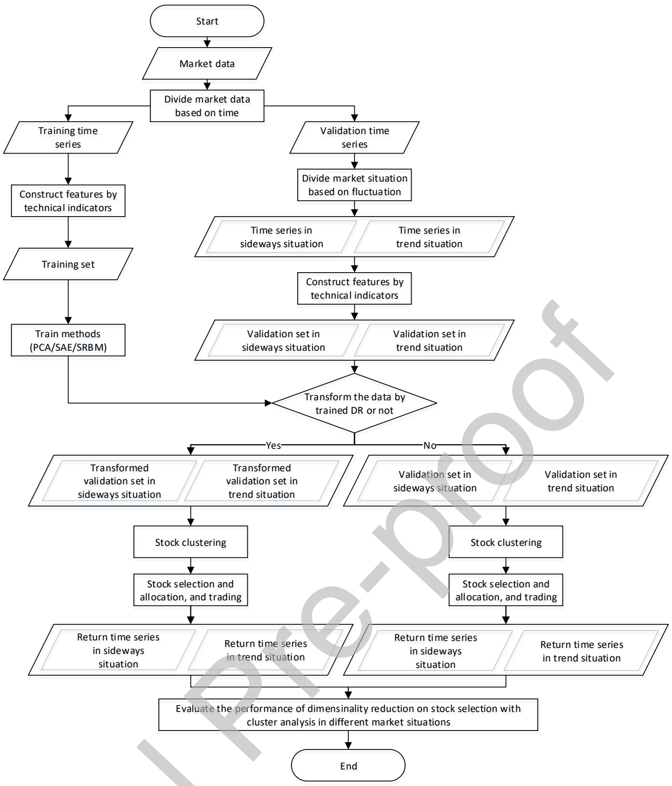


Fig. 1. Decision pipeline of the effect of dimensionality reduction on stock selection with cluster analysis in trend and sideways situations. Dimensionality-reduction methods include principal component analysis (PCA), stacked autoencoder (SAE), and stacked restricted Boltzmann machine (SRBM).

2. Methodology

In this section, three dimensionality-reduction (DR) methods and a stock-selection strategy with cluster analysis are introduced. For DR methods, we choose principal component analysis (PCA), stacked autoencoder (SAE), and stacked restricted Boltzmann machine (SRBM). Among them, PCA is a conventional model, and SAE and SRBM are deep-learning models.

For convenience, we first introduce some notations. An $n \times D$ matrix $\mathbf{X}$ represents sample data. Its row $\mathbf{x}$ is a D-dimensional vector representing a sample, and its column is an n-dimensional vector representing a feature. The use of DR methods on the matrix $\mathbf{X}$ will produce a transformed $n \times d$ matrix $\mathbf{Y}$.

2.1. Principal component analysis

Principal component analysis (PCA) is an unsupervised linear method which is widely used to reduce the dimension of data (Jolliffe and Cadima, 2016). It preserves the statistical information (variance and covariance) of the data as much as possible by embedding the data in a low-dimensional linear space.

Assuming a linear mapping $D \times d$ matrix $\mathbf{U}$, we can use $\mathbf{U}$ to transform the original sample matrix $\mathbf{X}$ into the matrix $\mathbf{Y}$ by the transformation $\mathbf{Y} = \mathbf{XU}$. The covariance of this transformed sample data can be calculated as

$$\mathbf{Y}^T\mathbf{Y} = (\mathbf{XU})^T(\mathbf{XU}) = \mathbf{U}^T(\mathbf{X}^T\mathbf{X})\mathbf{U} = \mathbf{U}^T\mathrm{cov}(\mathbf{X})\mathbf{U}, \tag{1}$$

where $\mathrm{cov}(\mathbf{X})$ is the covariance matrix of the original sample data $\mathbf{X}$.

The purpose of PCA is to maximize the sample data covariance. So, the linear mapping $\mathbf{U}$ consists of the d first principal eigenvectors of the matrix $\mathrm{cov}(\mathbf{X})$ with zero-mean $\mathbf{X}$. The eigenvector can be calculated by

$$\mathrm{cov}(\mathbf{X})\mathbf{v} = \lambda\mathbf{v}, \tag{2}$$

where λ, $\mathbf{v}$ are the eigenvalue and eigenvector, respectively, of $\mathrm{cov}(\mathbf{X})$.

With this linear mapping, we can transform the original sample $\mathbf{X}$ with dimension D to the transformed data $\mathbf{Y}$ with dimension d according to the transformation $\mathbf{Y} = \mathbf{XU}$.

2.2. Stacked autoencoder

Stacked autoencoder (SAE) is a model of deep neural networks, which is initialized by autoencoder to minimize the reconstruction error. It is an unfolded structure composed of one input layer, one hidden layer, and one reconstructed layer, as shown in Fig. 2.

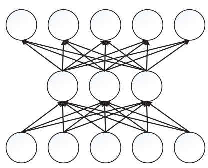


Fig. 2. The unfolded structure of the stacked autoencoder (SAE) with one hidden layer. The reconstruction error can be calculated by the difference between the input and reconstructed layer, and the model is trained by stochastic gradient descent (SGD) according to this error.

The updating of parameters of the SAE with one hidden layer is as follows. The vector $\mathbf{x}$ with dimension D is the input of the SAE. Assuming a weight matrix $\mathbf{W}$, bias vector $\mathbf{b}$ of the hidden layer, and bias vector $\mathbf{c}$ of the input layer, the reconstruction vector $\mathbf{x}^{'}$ is obtained as

$$\mathbf{x}^{'} = f(\mathbf{c} + \mathbf{W}\mathbf{y}), \tag{3}$$

$$\mathbf{y} = f(\mathbf{W}^T\mathbf{x} + \mathbf{b}), \tag{4}$$

where $f(\cdot)$ is the activation function and the vector $\mathbf{y}$ represents the transformed features of $\mathbf{x}$. We use the root-mean-square error (RMSE) to measure the reconstruction error or loss according to Eq. (5). The weight matrix $\mathbf{W}$ and two bias vectors $\mathbf{b}$, $\mathbf{c}$ can be adjusted by the back-propagation algorithm as

$$RMSE(\mathbf{x},\mathbf{x}^{'}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{'}\|^2, \tag{5}$$

$$\Delta\mathbf{W} = -\eta\frac{\partial}{\partial\mathbf{W}}RMSE(\mathbf{x},\mathbf{x}^{'}), \tag{6}$$

$$\Delta\mathbf{b} = -\eta\frac{\partial}{\partial\mathbf{b}}RMSE(\mathbf{x},\mathbf{x}^{'}), \tag{7}$$

$$\Delta\mathbf{c} = -\eta\frac{\partial}{\partial\mathbf{c}}RMSE(\mathbf{x},\mathbf{x}^{'}), \tag{8}$$

where η is the learning rate, and $\frac{\partial}{\partial\mathbf{W}}$, $\frac{\partial}{\partial\mathbf{b}}$, and $\frac{\partial}{\partial\mathbf{c}}$ are the partial derivatives of the error or loss function $RMSE(\mathbf{x},\mathbf{x}^{'})$ in terms of the quantities $\mathbf{W}$, $\mathbf{b}$, and $\mathbf{c}$, respectively. The training is over when either the reconstruction error is convergence or the back-propagation algorithm reaches its maximum number of iterations.

The SAE with l hidden layers can be trained layer-by-layer. Activities on the $(l-1)$th layer can be treated as the input of the lth layer, so all parameters of the SAE with l hidden layers can be obtained by training l autoencoders with one hidden layer. Fig. 3 shows this training process for the SAE with two hidden layers.

After the training process like in Fig. 3, the SAE is further fine-tuned by the back-propagation algorithm. The unit number of the last hidden layer represents the dimension of the transformed space, and transformed features can be generated by the transformation $f(\mathbf{W}^T\mathbf{x}+\mathbf{b})$, where the vector $\mathbf{x}$ represents input features of the last hidden layer.

2.3. Stacked restricted Boltzmann machine

Stacked restricted Boltzmann machine (SRBM) is also a model of deep neural networks. It has the same structure as the stacked autoencoder (SAE), but a different training algorithm. The SRBM with one hidden layer is a restricted Boltzmann machine (RBM), whose structure is presented in Fig. 4.

Restricted Boltzmann machine (RBM) has visible and hidden units, and is an energy-based stochastic recurrent neural network model. The vectors $\mathbf{v}$ and $\mathbf{h}$ respectively denote the state of visible and hidden units. The vectors $\mathbf{v}^{'}$ and $\mathbf{h}^{'}$ represent the expectation state of visible and hidden units after Gibbs sampling, respectively. $\mathbf{W}$, $\mathbf{b}$ and $\mathbf{c}$ are the weight matrix, bias vectors of hidden and visible layers, respectively. $f(\cdot)$ is the sigmoid activation function. The training process of the RBM can be described as follows.

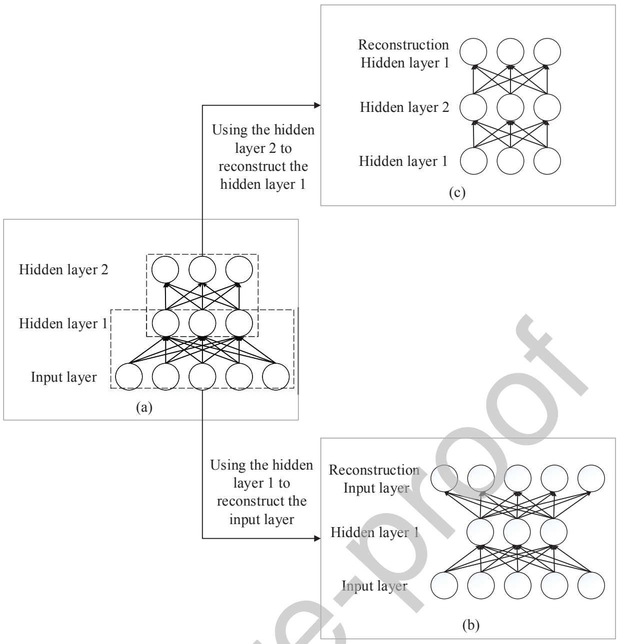


Fig. 3. The stacked autoencoder (SAE) with two hidden layers and the layer-by-layer training strategy. (a) Structure of the SAE with two hidden layers. (b) Unfolded structure of this SAE with the first hidden layer. (c) Unfolded structure of this SAE with the second hidden layer. We train two autoencoders layer-by-layer for this SAE. The unit number of the second hidden layer represents the dimension of the transformed space.

- The joint probability between visible and hidden units is expressed as

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{S} \mathrm{e}^{E(\mathbf{v}, \mathbf{h})}, \tag{9}$$

$$E(\mathbf{v}, \mathbf{h}) = -(\mathbf{v}^T \mathbf{W} \mathbf{h} + \mathbf{b}^T \mathbf{h} + \mathbf{c}^T \mathbf{v}), \tag{10}$$

  where $E(\mathbf{v}, \mathbf{h})$ is the energy function, and S is the partition function to assure that the probabilities sum up to 1.

- The marginal probability of visible units can be calculated as

$$P(\mathbf{v}) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}). \tag{11}$$

- The weight matrix is updated by the contrastive divergence (CD) algorithm as

$$\Delta \mathbf{W} = \eta \frac{\partial \ln(P(\mathbf{v}))}{\partial \mathbf{W}} = \eta(\mathbf{v}\mathbf{h}^T - \mathbf{v}^{'} \mathbf{h}^{'T}), \tag{12}$$

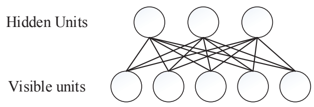


Fig. 4. The restricted Boltzmann machine (RBM).

where η is the learning rate.

- The bias vectors of visible and hidden units are also updated by the CD algorithm as

$$\Delta \mathbf{b} = \eta(\mathbf{v} - \mathbf{v}^{'}), \tag{13}$$

$$\Delta \mathbf{c} = \eta(\mathbf{h} - \mathbf{h}^{'}). \tag{14}$$

Stacked restricted Boltzmann machine (SRBM) can also be trained layer-by-layer. The transformed features can be generated by the transformation $f(\mathbf{W}^T\mathbf{v} + \mathbf{b})$, where the vector $\mathbf{v}$ represents the state of visible units in the last hidden layer and the number of visible units represents the dimension of the transformed space.

2.4. A stock-selection strategy with cluster analysis

Cluster analysis is a frequently used method and is crucial to stock selection. In this study, we analyze stock clusters by the affinity propagation (AP) algorithm (Frey and Delbert, 2007), whose cluster numbers does not be prespecified and results are not affected by random seeds. The process of AP algorithm is: (1) initializing the availabilities $a(i,k)$ to zero and the responsibilities $r(i,k)$ to the input similarity between objects i and k; (2) updating all responsibilities, given the availabilities, by Eq. 15; (3) updating all availabilities, given the responsibilities, by Eq. 16; (4) repeating steps (2) and (3), and combining availabilities and responsibilities to monitor the exemplar decisions and terminate the algorithm when these decisions do not change for a given number of iterations.

$$r(i,k) = s(i,k) - \max_{k^{'} \neq k}\{a(i,k^{'}), s(i,k^{'})\}, \tag{15}$$

$$a(i,k) = \min\{0, r(k,k) + \sum_{i^{'} \notin i,k}\{0, r(i^{'},k)\}\}, \tag{16}$$

where $r(i,k)$ and $a(i,k)$ represent the responsibility sent from object i to exemplar k and the availability sent from exemplar k to object i. $s(i,k)$ represents the similarity between objects i and k.

With the clustered stocks, we construct our portfolio and set the trading strategy. In terms of stock selection, as we all know, select diverse stocks can reduce the systemic risk of portfolios, so we directly select one stock with the highest Sharpe ratio in the past from each stock cluster and set an equal allocation for them, as in Dary et al. (2013), Plyakha et al. (2014) and Hu et al. (2018). For trading, we re-select and reallocate stocks each week with the transaction tax set to 0.0004. By the back-testing over a period, we obtain the return time series of this stock selection, and eventually use the Sharpe ratio to evaluate the performance of different stock-selection strategies.

3. The effect of dimensionality reduction on stock selection with cluster analysis in different market situations

In this section, we explore the effect of dimensionality reduction on stock selection with cluster analysis in different situations. Firstly, we train dimensionality-reduction methods in the training set. Then, we employ the trained dimensionality-reduction methods to reduce stock characteristics in the validation set. Finally, in the validation set, we compare the performance between the stock selection with and without dimensionality reduction in different market situations.

3.1. The training of dimensionality-reduction methods

At first, the China Securities 100 Index (CSI 100) and its constituent stocks are used to introduce the construction of input characteristics and analyze the training time and error for dimensionality-reduction methods. A neural network trained by all stocks represents index characters (Heaton et al., 2017). To reduce the training time for stacked autoencoder (SAE) and stacked restricted Boltzmann machine (SRBM), we apply the CSI 100 to train three dimensionality-reduction methods directly. All original weekly data of the CSI 100 (Date, Open, High, Low, Close) are downloaded from the CHOICE database (http://stock.eastmoney.com/). Data from Jun. 2, 2006, to Dec. 27, 2013, are used to train dimensionality-reduction methods and shown in Fig. 5.

Then, according to the open, high, low, and close of the CSI 100, eight frequently used technical indicators (Commodity Channel Index, CCI; Momentum, MOM; Moving Average Convergence Divergence, MACD; Relative Strength Index, RSI; Williams′%R, WillR; Simple Moving Average, SMA; Stochastic %K, StochK; Stochastic %D, StochD) are adopted to characterize stocks. The parameter settings of those indicators are in Table 1. These technical indicators calculated by TA-lib (http://ta-lib.org/) are shown in Fig. 6.

Table 1
The parameters of technical indicators.

Technical	Parameters			
Indicators	Time Period	Fast Period	Slow Period	Signal Period
CCI	5	-	-	-
MOM	5	-	-	-
MACD	-	3	5	10
RSI	5	-	-	-
WillR	5	-	-	-
SMA	5	-	-	-
StochK	-	-	-	-
StochD	-	-	-	-

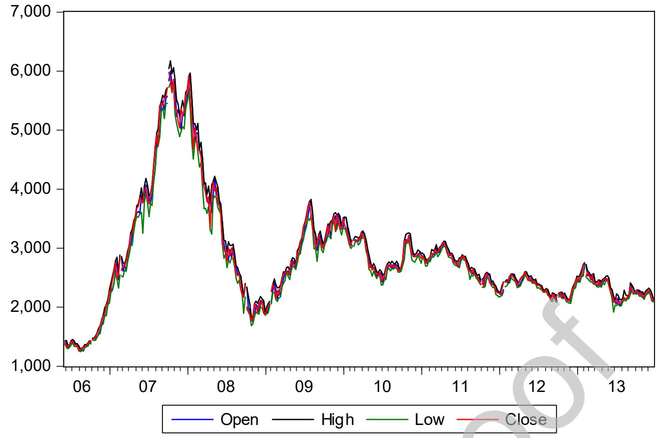


Fig. 5. The weekly data of open, high, low, and close of the CSI 100.

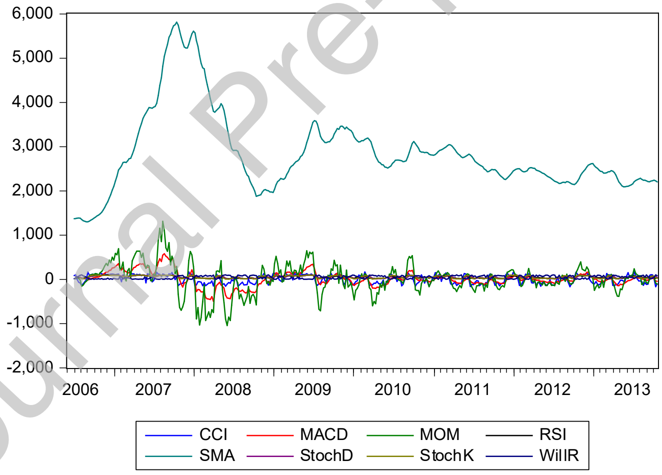


Fig. 6. The weekly technical indicators of the CSI 100.

Technical indicators are then processed into trend-deterministic data because they have better deterministic performance than continuous data in stock markets (Patel et al., 2015). Deterministic rules and weekly trend-deterministic indicators are shown in Table 2 and Fig. 7, respectively. As can be seen from Fig. 7, there are obvious differences among the eight weekly trend-deterministic indicators. For example, compared with other indicators, the changes of momentum and simple moving average are relatively slow, which means the trend of these indicators maybe continue for a while. In other words, momentum and simple moving average are not time-sensitive.

Table 2
The rules of trend-deterministic indicators. CCI(-1) represents the value of CCI in the last period, and the others are similar to CCI(-1).

Technical Indicators	Rules 1	Rules 0
CCI	>200 or >CCI(-1)	<-200 or <CCI(-1)
MOM	>0	<0
MACD	>MACD(-1)	<MACD(-1)
RSI	<30 or >RSI(-1)	>70 or <RSI(-1)
WillR	>WillR(-1)	<WillR(-1)
SMA	>Close	<Close
StochK	>StochK(-1)	<StochK(-1)
StochD	>StochD(-1)	<StochD(-1)

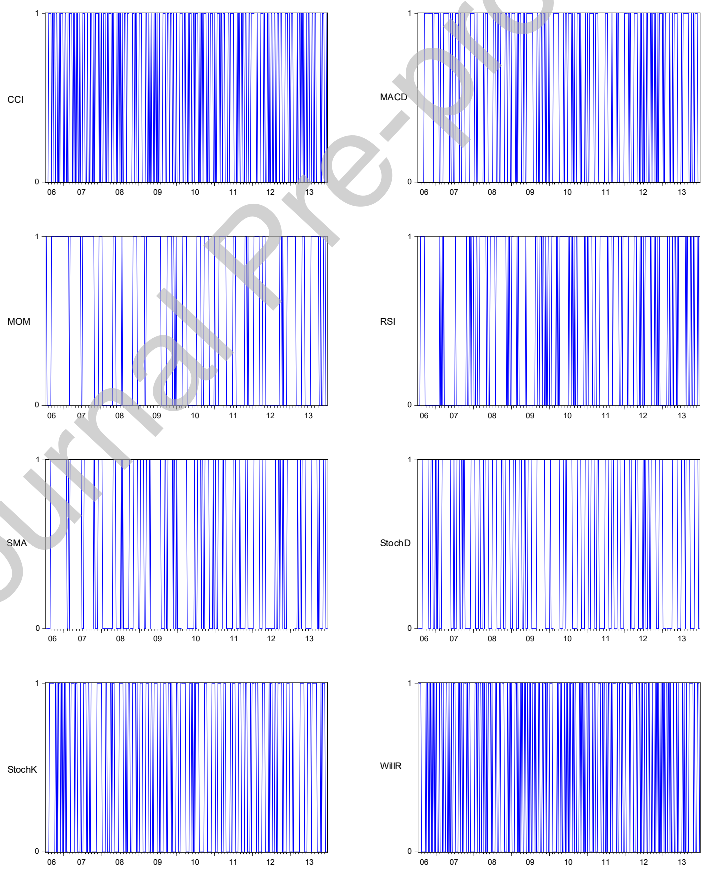


Fig. 7. The weekly trend-deterministic indicators of the CSI 100.

The input characteristics are the trend-deterministic indicators in the last eight weeks (two months), which means there are 64 input characteristics for dimensionality-reduction methods. Table 3 shows an example of input characteristics in the training set.

Table 3
An example of input characteristics of dimensionality-reduction methods.

	2013/12/27							
	CCI	MOM	MACD	RSI	WillR	SMA	StochK	StochD
2013/11/01	1	1	1	1	0	0	1	0
2013/11/08	0	0	0	0	1	0	0	0
2013/11/15	1	0	1	1	0	0	1	1
2013/11/22	1	0	1	1	0	1	1	1
2013/11/29	1	1	1	1	0	1	1	1
2013/12/06	0	1	1	1	1	1	1	1
2013/12/13	0	1	0	0	1	0	0	1
2013/12/20	0	0	0	1	1	0	0	0

Parameter configuration has a significant impact on different dimensionality-reduction methods (Bengio, 2012; Hinton, 2012). The most important parameter of principal component analysis (PCA) is the number of principal components. Parameters of stacked autoencoder (SAE) and stacked restricted Boltzmann machine (SRBM) mainly include the number of hidden layers, unit number of each layer, activation function, learning rate, number of training epochs, and batch size. According to the practical guide in Bengio (2012) and Hinton (2012), the configuration of these parameters is described in Table 4.

Table 4
The parameter configuration of dimensionality-reduction methods. In this table, [10,60,5] represents the numbers from 10 to 60 with interval 5. SGD represents the stochastic gradient descent algorithm, and CG is the contrast gradient algorithm.

Parameters	PCA	SAE	SRBM
Principal components	[10,60,5]	-	-
Number of hidden-layer neurons	-	[10,60,5]	[10,60,5]
Number of hidden layers	-	1	1
Activation function	-	sigmoid	sigmoid
Training algorithm	-	SGD	CG
Gibbs sampling k-steps	-	-	1
Training epochs	-	100	100
Learning rate	-	0.001	0.001
Training batch size	-	4	4

According to the methods in section 2, the input characteristics of PCA must be zero-mean, so all training data of PCA are standardized by Z-standardization. The input characteristics of SRBM and SAE are the trend-deterministic data. To avoid overfitting for training SRBM and SAE, 20% of the training data are used to monitor the training error and the others are used to train them. When the moving average of the last five training errors stops decreasing, the training processes of SRBM and SAE is terminated. For different dimensionality-reduction methods, Table 5 shows an example of reduced characteristics, which corresponds to the input characteristics in Table 3. It can be seen from Table 5 that the reduced characteristics of SRMB and SAE are larger than zero, which is caused by the sigmoid activation function with the value between 0 and 1.

However, there are both positive and negative values for PCA since the input characteristics of PCA must be zero-mean.

Table 5
An example of reduced characteristics whose dimension is 25.

	2013/12/27
PCA	–0.34, –3.25, +2.97, –2.76, –3.29, –1.46, –0.10, +0.70, +1.23, –1.08, –0.20, –1.47, –1.80, –0.71, –0.21, +0.65, +0.83, –1.12, +0.96, –0.51, –0.47, –0.89, –0.97, +0.36, +0.21.
SAE	+0.08, +0.15, +0.95, +0.65, +0.03, +0.49, +0.99, +0.50, +0.90, +0.84, +0.42, +0.58, +0.29, +0.40, +0.87, +1.00, +0.01, +0.02, +0.01, +0.99, +1.00, +0.00, +0.31, +0.98, +0.00.
SRBM	+0.97, +0.15, +0.03, +0.10, +0.15, +0.34, +0.83, +0.10, +0.11, +0.55, +0.42, +0.07, +0.97, +0.53, +0.72, +0.15, +0.63, +0.04, +0.04, +0.90, +0.01, +0.37, +0.32, +0.78, +0.05.

Table 6 shows the training time and error of different dimensionality-reduction methods in the training set. There are no obvious relations between dimensions and training time for the three dimensionality-reduction methods, but the training error decreases with dimensions. Surprisingly, PCA is much more efficient than SAE and SRBM. Because training mechanisms of the three dimensionality-reduction methods are differ, we do not compare the training error among them.

Table 6
Training time and error of PCA, SAE, and SRBM for the CSI 100. The training time is measured in seconds. The training error of PCA is the unexplained variance ratio, while SAE and SRBM are the mean square error.

Dimension	Training times (s)			Training error		
	PCA	SAE	SRBM	PCA	SAE	SRBM
10	0.050	11.930	10.273	0.300	0.136	0.216
15	0.040	11.559	13.908	0.237	0.116	0.186
20	0.000	15.104	10.603	0.188	0.086	0.191
25	0.000	13.823	12.354	0.150	0.080	0.178
30	0.000	12.376	9.923	0.117	0.080	0.183
35	0.000	13.856	10.608	0.090	0.068	0.174
40	0.000	16.922	9.543	0.067	0.056	0.177
45	0.000	11.694	9.271	0.047	0.067	0.177
50	0.000	11.979	11.513	0.031	0.063	0.166
55	0.002	13.251	10.157	0.018	0.055	0.165
60	0.000	14.553	10.987	0.007	0.052	0.162
avg	0.006	13.368	10.831	0.114	0.078	0.179

3.2. The effect of dimensionality reduction on stock selection with cluster analysis

Here, four stock-selection strategies and the indices are examined. They are stock selection with principal component analysis and cluster analysis (DR-CA-SS_{pca}), stock selection with stacked autoencoder and cluster analysis (DR-CA-SS_{sae}), stock selection with stacked restricted Boltzmann machine and cluster analysis (DR-CA-SS_{srbm}), stock selection with cluster analysis (CA-SS), and the indices (IND). CA-SS and IND are the benchmarks in our work because CA-SS uses no dimensionality reduction and IND is the indices of constituent stocks.

The CSI 100 constituent stocks, ranging from Jan. 3, 2014, to Feb. 26, 2016, are firstly used as the validation set to explore the effect of dimensionality reduction on stock selection. The construction of characteristics for each stock is similar to that in subsection 3.1. The dimensionality-reduction method and Z-standardization obtained in subsection 3.1 are applied to reduce the dimension of stock characteristics in the validation set. These unreduced and reduced trend-deterministic data (see Table 3 and 5) are applied to analyze the stock cluster and selection.

For stock selection strategies, the CSI 100 constituent stocks, whose characteristics are generated from the last eight weeks, are firstly clustered by the affinity propagation (AP) algorithm each week. Then, stocks with the highest Sharpe ratio in the last eight weeks are selected from each stock cluster. All selected stocks are re-selected and reallocated in equal proportions each week. An example of selected stocks of DR-CA-SS$_{pca}$ and CA-SS is illustrated in Table 7.

Table 7
An example of selected stocks of DR-CA-SS$_{pca}$ and CA-SS for the CSI 100 constituent stocks on Nov. 7, 2014.

Stock Selection	Dimension	Selected Stocks	Return
DR-CA-SS$_{pca}$	25	601800.SH, 601988.SH, 300059.SZ 601009.SH, 600900.SH, 601998.SH 601336.SH, 601618.SH, 000001.SZ 002450.SZ, 600015.SH	0.0368
	50	601800.SH, 600837.SH, 601288.SH 601009.SH, 601998.SH, 601336.SH 601668.SH, 601618.SH, 000001.SZ 002450.SZ, 600015.SH	0.0327
CA-SS	64	601800.SH, 600036.SH, 601009.SH 002415.SZ, 601998.SH, 600900.SH 601336.SH, 601988.SH, 601390.SH 002450.SZ, 600015.SH	0.0213

After stock selection and trading, weekly returns of different strategies are obtained from Jan. 3, 2014, to Feb. 26, 2016, so the Sharpe ratio is directly used to evaluate the performance of these strategies. The weekly Sharpe ratio SR, which consists of the return and risk, can be calculated as

$$SR = \frac{E(\boldsymbol{r}) - \hat{r}}{\sigma(\boldsymbol{r})}, \tag{17}$$

where the vector $\boldsymbol{r} = [r_{t_o}, r_{t_{o+1}}, ..., r_t, \cdots, r_{t_{c-1}}, r_{t_c}]$ represents weekly returns of strategies, and the quantity r_t is the return of strategies in the tth week. The quantities $E(\boldsymbol{r})$, $\sigma(\boldsymbol{r})$ denote the mean and standard deviation of vector $\boldsymbol{r}$, respectively. The quantity $\hat{r}$ represents the risk-free return, which is set to zero in this work.

Fluctuation is an important consideration in identifying market situations (Hanna, 2018). Following Kirkpatrick and Dahlquist (2010), we divide the market into sideways and trend situations, where the trend situation includes up and down. The dividing result of different situations is shown in Fig. 8.

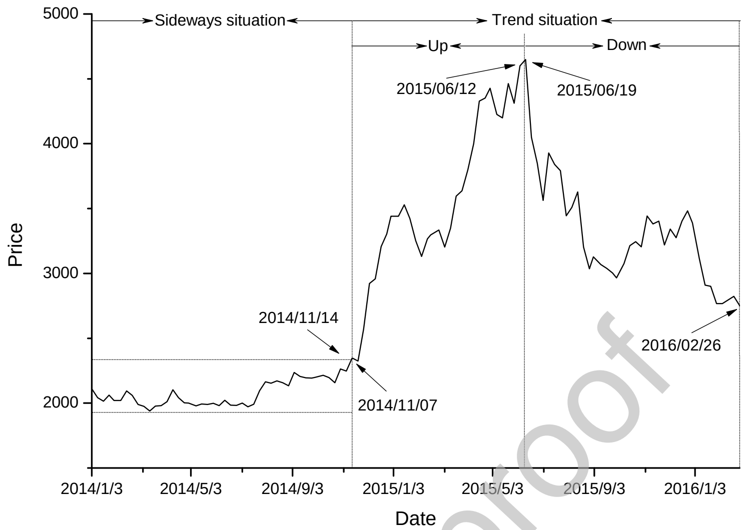


Fig. 8. The different situations of the CSI 100.

Table 8 shows the detailed Sharpe ratios of stock-selection strategies for the CSI 100 constituent stocks. It can be seen that the dimensions of the best Sharpe ratios in the validation set are 15, 15, and 25 for PCA, SAE, and SRBM, respectively. In particular, these Sharpe ratios fluctuate with dimensions or training error, which means there are no obvious relations between dimensions and performance of stock selection in the validation set. That is to say, we cannot use the best dimension to analyze the effect of dimensionality reduction on stock selection, so we summarize the Sharpe ratios in dimensions, as shown in Table 9.

According to Table 9, we get the following conclusions: (1) in the sideways situation, both PCA and SAE decrease the Sharpe ratio of CA-SS, while SRBM slightly improves the Sharpe ratio of CA-SS; (2) in the up-trend situation, the Sharpe ratios of stock selection with the three dimensionality-reduction methods show no obvious difference from CA-SS; (3) in the down-trend situation, the dimensionality reduction can significantly improve the Sharpe ratio of CA-SS. Some factors may contribute to these results. Firstly, compared to mathematical data, financial data carry a lower signal-to-noise ratio. Although dimensionality reduction loses part of information, it preserves the main information and avoids the curse of dimensionality, so it can improve the Sharpe ratio of stock selection with cluster analysis in some situations. Secondly, the effect of dimensionality reduction in different situations may depend on the market analyzed. For example, the decline of the CSI 100 can easily trigger investor panic, especially for noise traders, which makes relations among stock characteristics complicated and noisy. Therefore, the signal-to-noise ratio of the data can be improved by

Table 8
The detailed Sharpe ratios of stock-selection strategies for the CSI 100 constituent stocks in different situations. There is a sideways situation from Jan. 3, 2014, to Nov. 7, 2014, and a trend situation from Nov. 14, 2014, to Feb. 26, 2016. "All" represents the all validation set, which is from Jan. 3, 2014, to Feb. 26, 2016. The average Sharpe ratio with 30 times is used to reduce the effect of random seeds on DR-CA-SS$_{sae}$ and DR-CA-SS$_{srbm}$.

Stock Selection	Dimension	Sideways	Trend		All
			Up	Down	
DR-CA-SS$_{pca}$	10	0.1768	0.6675	-0.2113	0.1400
	15	0.1639	0.6515	-0.1836	**0.1537**
	20	0.1646	0.6419	-0.1587	0.1511
	25	0.1460	0.6238	-0.1786	0.1222
	30	0.1218	0.6540	-0.1536	0.1405
	35	0.1649	0.6485	-0.1901	0.1309
	40	0.1035	0.6189	-0.1652	0.1072
	45	0.1258	0.6128	-0.2099	0.1018
	50	0.1449	0.6180	-0.2157	0.1077
	55	0.1593	0.5432	-0.2411	0.0870
	60	0.1563	0.6299	-0.2169	0.1245
DR-CA-SS$_{sae}$	10	0.1533	0.5993	-0.2198	0.1121
	15	0.1653	0.6266	-0.2088	**0.1231**
	20	0.1606	0.6236	-0.2077	0.1231
	25	0.1763	0.6160	-0.2131	0.1220
	30	0.1290	0.6221	-0.2125	0.1150
	35	0.1447	0.6126	-0.2071	0.1180
	40	0.1513	0.6152	-0.2127	0.1160
	45	0.1446	0.6035	-0.2000	0.1194
	50	0.1377	0.6308	-0.2130	0.1176
	55	0.1361	0.6120	-0.1991	0.1197
	60	0.1440	0.6161	-0.2029	0.1211
DR-CA-SS$_{srbm}$	10	0.1596	0.6228	-0.1740	0.1452
	15	0.1786	0.6413	-0.1695	0.1571
	20	0.1841	0.6251	-0.1587	0.1555
	25	0.2128	0.6234	-0.1632	**0.1592**
	30	0.2174	0.6365	-0.1786	0.1542
	35	0.2141	0.6295	-0.1722	0.1550
	40	0.2310	0.6406	-0.1795	0.1575
	45	0.2269	0.6332	-0.1776	0.1534
	50	0.2167	0.6412	-0.1843	0.1461
	55	0.2375	0.6344	-0.1815	0.1515
	60	0.2163	0.6324	-0.1854	0.1440
CA-SS	64	0.1923	0.6241	-0.2349	0.1221
IND	–	0.0624	0.5377	-0.2625	0.0744

dimensionality reduction, thereby improving the performance of stock selection in this decline of the market. Especially, considering that noise traders make markets fluctuate (Verma and Verma, 2007), we speculate that the advantage of dimensionality reduction mainly appears in trend situations, but whether it is in an up or down trend may depend on the market analyzed.

To verify the findings above, we further analyze different strategies for the Nikkei 225 constituent stocks. Similar to our analysis for the CSI 100, we use the data from Jan. 1, 2000, to Dec. 27, 2013, to train DR methods, and the data from Jan. 5, 2014, to Feb. 14, 2016, to evaluate and compare the performance of different strategies. The dividing result of different market situations is shown in Fig. 9.

Table 10 shows the detailed Sharpe ratios of stock-selection strategies for the Nikkei 225 constituent stocks. The conclusions obtained from Table 10

Table 9
Sharpe ratio summary of stock-selection strategies in dimensions for the CSI 100 constituent stocks. ***, **, and * denote the significance of t-test at 1%, 5%, and 10%, respectively, in dimensions between stock selection with and without dimensionality reduction.

Situations	Stock Selection	Min	Max	Mean	Std
Sideways	DR-CA-SS$_{pca}$***	0.1035	0.1768	0.1480	0.0224
	DR-CA-SS$_{sae}$***	0.1290	0.1763	0.1494	0.0139
	DR-CA-SS$_{srbm}$**	0.1596	0.2375	0.2086	0.0241
	CA-SS	0.1923	0.1923	0.1923	0
	IND	0.0624	0.0624	0.0624	0
Trend(Up)	DR-CA-SS$_{pca}$	0.5432	0.6675	0.6282	0.0332
	DR-CA-SS$_{sae}$***	0.5993	0.6308	0.6162	0.0095
	DR-CA-SS$_{srbm}$***	0.6228	0.6413	0.6328	0.0069
	CA-SS	0.6241	0.6241	0.6241	0
	IND	0.5377	0.5377	0.5377	0
Trend(Down)	DR-CA-SS$_{pca}$***	-0.2411	-0.1536	-0.1959	0.0264
	DR-CA-SS$_{sae}$***	-0.2198	-0.1991	-0.2088	0.0063
	DR-CA-SS$_{srbm}$***	-0.1854	-0.1587	-0.1750	0.0085
	CA-SS	-0.2349	-0.2349	-0.2349	0
	IND	-0.2625	-0.2625	-0.2625	0

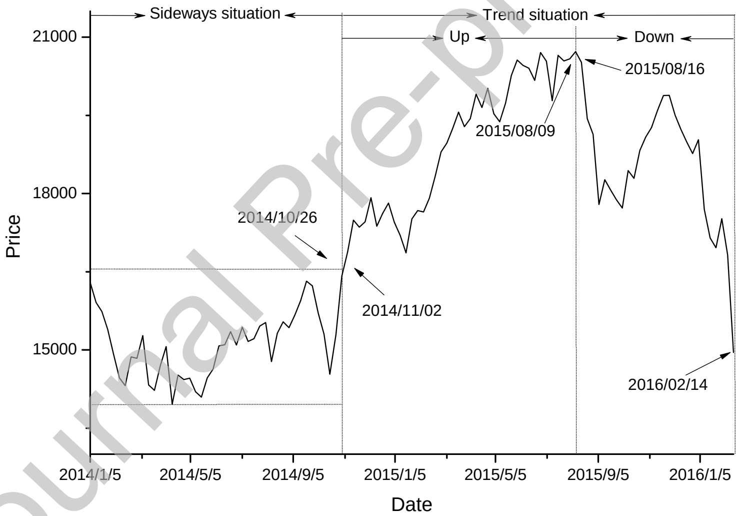


Fig. 9. The different situations of the Nikkei 225.

are in agreement with those in Table 8. That is to say, there are also no obvious relations between dimensions and performance of stock selection in the validation set.

Table 11 shows the summarized Sharpe ratios of stock-selection strategies for the Nikkei 225 constituent stocks. In the sideways situation, both PCA and SAE decrease the Sharpe ratio of CA-SS. Although SRBM improves the Sharpe ratio of CA-SS, the improvement is minimal and is not statistically significant. In the up-trend situation, all the three dimensionality-reduction methods can significantly improve the Sharpe ratio of CA-SS, both statistically and quantita-

Table 10
The detailed Sharpe ratios of stock-selection strategies for the Nikkei 225 constituent stocks in different situations. It is a sideways situation from Jan. 5, 2014, to Oct. 26, 2014, and a trend situation from Nov. 2, 2014, to Feb. 14, 2016. "All" represents the all validation set, which is from Jan. 5, 2014, to Feb. 14, 2016. The average Sharpe ratio with 30 times is used to reduce the effect of random seeds on DR-CA-SS$_{sae}$ and DR-CA-SS$_{srbm}$.

Stock Selection	Dimension	Sideways	Trend		All
			Up	Down	
DR-CA-SS$_{pca}$	10	-0.022	0.4307	-0.2968	0.0290
	15	0.0155	0.4688	-0.3453	0.0370
	20	-0.0068	0.4631	-0.3380	0.0236
	25	-0.0212	0.4402	-0.3147	0.0190
	30	0.0076	0.4414	-0.2838	0.0461
	35	0.0140	0.3905	-0.2815	0.0428
	40	0.0108	0.4491	-0.3071	0.0518
	45	0.0108	0.4682	-0.3125	**0.0529**
	50	-0.0108	0.4293	-0.3011	0.0320
	55	-0.0142	0.4000	-0.3417	0.0224
	60	-0.0488	0.3910	-0.3129	0.0144
DR-CA-SS$_{sae}$	10	-0.0163	0.3841	-0.3422	0.0081
	15	-0.0160	0.4103	-0.3283	**0.0133**
	20	-0.0188	0.4076	-0.3335	0.0113
	25	-0.0194	0.4003	-0.3348	0.0075
	30	-0.0222	0.4030	-0.3369	0.0086
	35	-0.0260	0.4068	-0.3406	0.0073
	40	-0.0245	0.4147	-0.3341	0.0107
	45	-0.0184	0.3934	-0.3406	0.0069
	50	-0.0281	0.3927	-0.3452	0.0024
	55	-0.0151	0.3999	-0.3389	0.0115
	60	-0.0244	0.4026	-0.3448	0.0058
DR-CA-SS$_{srbm}$	10	0.0112	0.4782	-0.4025	0.0334
	15	0.0172	0.4715	-0.3871	**0.0398**
	20	0.0008	0.4551	-0.3963	0.0266
	25	0.0036	0.4566	-0.3754	0.0323
	30	0.0104	0.4524	-0.3763	0.0340
	35	0.0035	0.4341	-0.3779	0.0262
	40	0.0027	0.4387	-0.3751	0.0273
	45	-0.0039	0.4430	-0.3633	0.0287
	50	-0.0028	0.4481	-0.3627	0.0314
	55	-0.0092	0.4403	-0.3866	0.0214
	60	-0.0090	0.4402	-0.3582	0.0297
CA-SS	64	0.0004	0.3769	-0.3065	0.0283
IND	–	-0.0354	0.3494	-0.3312	-0.0117

tively. In the down-trend situation, all three dimensionality-reduction methods decrease the Sharpe ratios of CA-SS. Therefore, the advantage of dimensionality reduction is reflected in the trend situation. However, unlike the CSI 100 constituent stocks, dimensionality reduction significantly improves the performance of stock selection in the up-trend situation, which means the advantage of dimensionality reduction depends on the market analyzed.

In sum, from a series of experiments, we get the following three conclusions. Firstly, although SRBM and SAE can learn nonlinear relations among characteristics, they are prone to overfitting and their learning efficiency is low. More importantly, as non-parametric methods, they are susceptible to random seeds. That is to say, SRBM and SAE have no advantages over PCA except for fitting nonliear relations. Secondly, for dimensionality-reduction methods, the optimal dimension in the training set is not necessarily optimal in the validation set,

Table 11
Sharpe ratio summary of stock-selection strategies in dimensions for the Nikkei 225 constituent stocks. ***, **, and * denote the significance of t-test at 1%, 5%, and 10%, respectively, in dimensions between stock selection with and without dimensionality reduction.

Situations	Stock Selection	Min	Max	Mean	Std
Sideways	DR-CA-SS$_{pca}$	-0.0488	0.0155	-0.0059	0.0201
	DR-CA-SS$_{sae}$***	-0.0281	-0.0151	-0.0208	0.0044
	DR-CA-SS$_{srbm}$	-0.0092	0.0172	0.0022	0.0084
	CA-SS	0.0004	0.0004	0.0004	0
	IND	-0.0355	-0.0355	-0.0355	0
Trend(Up)	DR-CA-SS$_{pca}$***	0.3905	0.4688	0.4338	0.0291
	DR-CA-SS$_{sae}$***	0.3841	0.4147	0.4014	0.0088
	DR-CA-SS$_{srbm}$***	0.4341	0.4782	0.4507	0.0140
	CA-SS	0.3769	0.3769	0.3769	0
	IND	0.3494	0.3494	0.3494	0
Trend(Down)	DR-CA-SS$_{pca}$	-0.3453	-0.2815	-0.3123	0.0219
	DR-CA-SS$_{sae}$***	-0.3452	-0.3283	-0.3382	0.0052
	DR-CA-SS$_{srbm}$***	-0.4025	-0.3582	-0.3783	0.0140
	CA-SS	-0.3065	-0.3065	-0.3065	0
	IND	-0.3312	-0.3312	-0.3312	0

perhaps due to frequent changes in the stock market. Thirdly, the advantage of dimensionality reduction is mainly reflected in the trend situation, but whether it is in an up or down trend depends on the market analyzed. For the CSI 100 market, dimensionality reduction significantly improves the performance of stock selection in down trends, while for the Nikkei 225 market it significantly improves the performance in up trends.

4. A stock-selection rotation strategy based on the effect of dimensionality reduction

4.1. A stock-selection rotation strategy

Based on the effect of dimensionality reduction on stock selection in different market situations explored in section 3 and assuming that this effect will continue, we propose a stock-selection rotation strategy between stock selection with dimensionality reduction and cluster analysis (DR-CA-SS) and stock selection with cluster analysis (CA-SS), namely DR-CA-RSS. Firstly, we evaluate the significance of relations between the DR-CA-SS and CA-SS by the $t-$test in dimensions. Then, if DR-CA-SS significantly outperforms CA-SS, which means that dimensionality reduction can improve the performance of stock selection, we utilize DR-CA-SS. Otherwise, we continue to employ CA-SS. The pseudocode for the proposed stock-selection rotation strategy is shown in **Algorithm 1**.

Algorithm 1 A stock-selection rotation strategy between DR-CA-SS and CA-SS, namely DR-CA-RSS.

Input: initial investment strategy, formation period FT, and readjusting period RT, the date of open position t_o, the date of close position t_c.

$t = t_o + RT$
while $t < t_c$ **do**
Calculate Sharpe ratios of DR-CA-SS and CA-SS in last FT weeks
Evaluate the significance of $t-$test between DR-CA-SS and CA-SS in dimensions
    if The current investment strategy is CA-SS **then**
        if The performance of DR-CA-SS is **significantly larger** than that of CA-SS **then** Utilize DR-CA-SS
        else Continue utilizing CA-SS
        end if
    end if
    if The current investment strategy is DR-CA-SS **then**
        if The performance of DR-CA-SS is **(significantly) smaller** than that of CA-SS **then** Utilize CA-SS
        else Continue utilizing DR-CA-SS
        end if
    end if
$t = t + RT$
end while

4.2. Results

For the three dimensionality-reduction methods, considering that stacked autoencoder and stacked restricted Boltzmann machine are particularly sensitive to random seeds, which increases the investment risk, we only use principal component analysis (PCA) to implement the stock-selection rotation strategy, DR-CA-RSS$_{pca}$. The assessment criteria illustrated in subsection 3.2, Sharpe ratio, is used to evaluate the performance of this strategy.

Firstly, we test DR-CA-RSS$_{pca}$ for the CSI 100 constituent stocks in two different test periods. The first period spans from Mar. 4, 2016, to Aug. 24, 2018, which have never been used in our previous experiments. The second period is covered from Jan. 3, 2014, to Aug. 24, 2018, part of which has been used in the effect of dimensionality reduction on stock selection in different situations. Consistent with the experiment in section 3, the readjusting period, RT, is set to 1. The formation period, FT, is too small to calculate the Sharpe ratio. If it is too large, then the stock-selection rotation strategy becomes dull and loses trading opportunities. Therefore, the strategy is tested when FT is 3, 4, and 5. The initial investment strategy is set to CA-SS with better performance than the indices from the above experiments. Table 12 shows the detailed Sharpe ratios of different strategies in dimensions for the CSI 100 constituent stocks.

From Table 12, it can be seen that the average Sharpe ratio of DR-CA-RSS$_{pca}$ is higher than that of DR-CA-SS$_{pca}$, CA-SS, and the CSI 100 with regard to different test periods and formation periods. For example, when the FT is 4, average Sharpe ratios of DR-CA-RSS$_{pca}$ are 0.141 and 0.128 for two different test periods. From Jan. 3, 2014, to Aug. 24, 2018, average Sharpe ratios of DR-CA-RSS$_{pca}$ are 0.126, 0.128, and 0.129 for three different FT, and they are all higher than that of DR-CA-SS$_{pca}$ and CA-SS. What's more, compared with DR-CA-SS$_{pca}$, DR-CA-RSS$_{pca}$ has a higher significance with respect to CA-SS.

Table 12
Sharpe ratios of different strategies for the CSI 100 constituent stocks. p is the p value of the t-test between the strategy and CA-SS.

2016/03/04-2018/08/24														
	FT	Dimension											avg	p
		10	15	20	25	30	35	40	45	50	55	60		
DR-CA-RSS$_{pca}$	3	0.161	0.115	0.127	0.157	0.123	0.136	0.144	0.147	0.149	0.150	0.144	0.141	0.078
	4	0.157	0.121	0.135	0.150	0.128	0.139	0.139	0.143	0.140	0.151	0.143	0.141	0.028
	5	0.159	0.127	0.146	0.150	0.130	0.147	0.140	0.144	0.139	0.153	0.148	0.144	0.003
DR-CA-SS$_{pca}$	-	0.143	0.103	0.134	0.154	0.111	0.121	0.119	0.131	0.128	0.158	0.133	0.131	0.679
CA-SS	-						0.133							
IND	-						0.092							
2014/01/03-2018/08/24														
	FT	Dimension											avg	p
		10	15	20	25	30	35	40	45	50	55	60		
DR-CA-RSS$_{pca}$	3	0.133	0.120	0.135	0.138	0.130	0.123	0.120	0.116	0.120	0.124	0.130	0.126	0.078
	4	0.137	0.123	0.141	0.134	0.139	0.127	0.123	0.120	0.117	0.122	0.129	0.128	0.030
	5	0.130	0.128	0.140	0.128	0.143	0.132	0.129	0.121	0.117	0.116	0.131	0.129	0.025
DR-CA-SS$_{pca}$	-	0.137	0.129	0.139	0.131	0.124	0.123	0.109	0.110	0.112	0.111	0.124	0.123	0.834
CA-SS	-						0.122							
IND	-						0.078							

From Jan. 3, 2014, to Aug. 24, 2018, the p-values of the t-test between DR-CA-RSS$_{pca}$ and CA-SS are 0.078, 0.030, and 0.025, respectively, for different HT, and they are all significant. However, there is no significant difference between DR-CA-SS$_{pca}$ and CA-SS. Interestingly, there is a unanimous conclusion drawn from the experiments based on the data from Mar. 4, 2016, to Aug. 24, 2018. In addition, without loss of generality, we give an example of the cumulative return of different strategies from Jan. 3, 2014, to Aug. 24, 2018, in Fig. 10, when FT, RT, and the dimension of PCA are 5, 1, and 30, respectively. It can be obviously seen that DR-CA-RSS$_{pca}$ has a relatively higher cumulative return than that of DR-CA-SS$_{pca}$, CA-SS, and the CSI 100. That is to say, from the perspective of Sharpe ratios and the significance, the proposed strategy is better than DR-CA-SS$_{pca}$ and CA-SS.

To further demonstrate the effectiveness of the proposed strategy, we check other stock markets, including Shanghai Stock Exchange 180 (SSE 180), Nikkei 225, and S&P 500 constituent stocks. The data range from Feb. 21, 2016, to Aug. 26, 2018, for Nikkei 225 and S&P 500 and Feb. 19, 2016, to Aug. 24, 2018, for SSE 180. We provide Sharpe ratios of different strategies as shown in Table 13 when FT is 4, which means the formation period is one month.

As in the case for the CSI 100 constituent stocks, similar performance of the average Sharpe ratio can be obtained from Table 13. That is, the DR-CA-RSS$_{pca}$ brings more favorable Sharpe ratios than DR-CA-SS$_{pca}$ for each stock market. For example, the average Sharpe ratio of DR-CA-RSS$_{pca}$ is 0.151 for the SSE 180 constituent stocks, which is better than DR-CA-SS$_{pca}$, CA-SS, and the SSE 180. In conclusion, the above numerical results show that the stock-selection rotation strategy based on the effect of dimensionality reduction provides a valid and advantageous way to select stocks, and it's robust for many stock markets.

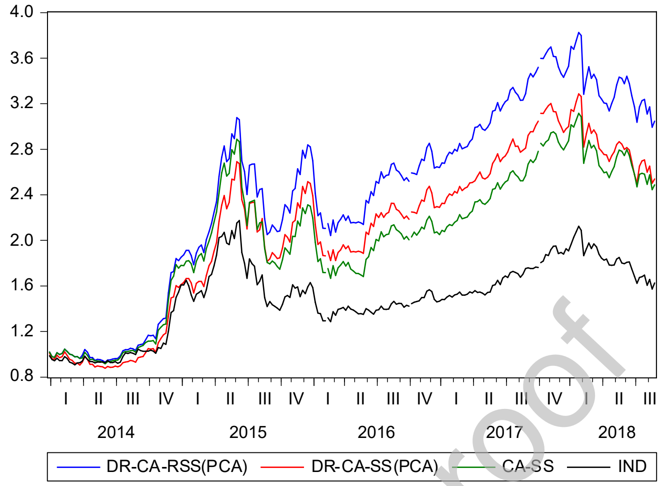


Fig. 10. An example of the cumulative return of different strategies from Jan. 3, 2014, to Aug. 24, 2018, for the CSI 100 constituent stocks.

Table 13
Sharpe ratios of different strategies for the SSE 180, Nikkei 225, and S&P 500 constituent stocks. It's the p value of the t-test between the strategy and CA-SS in parentheses.

Dimension	SSE 180		Nikkei 225		S&P 500	
	DR-CA-RSS$_{pca}$	DR-CA-SS$_{pca}$	DR-CA-RSS$_{pca}$	DR-CA-SS$_{pca}$	DR-CA-RSS$_{pca}$	DR-CA-SS$_{pca}$
10	0.135	0.148	0.141	0.094	0.286	0.268
15	0.153	0.161	0.148	0.115	0.237	0.253
20	0.144	0.161	0.156	0.143	0.226	0.239
25	0.147	0.159	0.157	0.159	0.251	0.265
30	0.148	0.151	0.166	0.172	0.255	0.268
35	0.151	0.136	0.147	0.130	0.252	0.231
40	0.155	0.141	0.148	0.141	0.204	0.273
45	0.176	0.168	0.156	0.137	0.248	0.223
50	0.149	0.132	0.153	0.120	0.224	0.202
55	0.152	0.131	0.154	0.122	0.227	0.216
60	0.154	0.141	0.140	0.126	0.208	0.211
avg	0.151 (0.000)	0.148 (0.001)	0.152 (0.040)	0.133 (0.023)	0.238 (0.002)	0.232 (0.031)
CA-SS	0.131		0.147		0.212	
IND	0.065		0.139		0.233	

5. Conclusions

Dimensionality reduction is an important process for stock selection with cluster analysis. It is not difficult to collect different kinds of data, hence the curse of dimensionality is inevitable in cluster analysis. Considering complex relations among dimensionality reduction, noise trading, and market situations, it is necessary to deeply understand the effect of dimensionality reduction on stock selection with cluster analysis in different market situations.

In this study, we first introduce three dimensionality reduction methods, including principal component analysis, stacked autoencoder, and stacked restricted Boltzmann machine, and present a stock-selection strategy with cluster analysis. Then, we analyze the effect of dimensionality reduction on stock selection with cluster analysis in sideways and trend situations, where the trend situation includes up and down, for the CSI 100 and Nikkei 225 constituent stocks. From a series of experiments, we find: (1) except for fitting nonlinear relations, stacked autoencoder and stacked restricted Boltzmann machine show no superiority to principal component analysis; (2) in sideways situations, dimensionality reduction hardly improves the performance of stock selection with cluster analysis; (3) the advantage of dimensionality reduction is mainly reflected in trend situations, but whether it is in an up or down trend depends on the market analyzed. For the CSI 100 constituent stocks, dimensionality reduction can significantly improve the performance of stock selection in down trends. While for the Nikkei 225 constituent stocks, dimensionality reduction can significantly improve the performance of stock selection in up trends. In addition, based on the empirical results, we propose a stock-selection rotation strategy between the stock selection with and without dimensionality reduction. The results of experiments show that the proposed rotation strategy outperforms the stock market indices as well as stock-selection strategies based on dimensionality reduction and cluster analysis. All these findings demonstrate both the superiority of this stock-selection rotation strategy and the importance of dimensionality reduction.

Ours is one of a few comprehensive studies to apply dimensionality reduction to stock selection. At first, we apply deep learning methods, including stacked autoencoder and stacked restricted Boltzmann machine, to finance. What's more, we analyze the effect of dimensionality reduction on stock selection with cluster analysis in different situations. Finally, we propose a stock-selection rotation strategy between stock selection with and without dimensionality reduction. This research can provide an important support for researchers and investors in dimensionality reduction and stock investment. However, there are still some limitations of our study, which presents opportunities for future research. For example, due to the capricious nature of stock markets, the effect of dimensionality reduction may change over time. Secondly, in section 4, because stacked restricted Boltzmann machine and stacked autoencoder are susceptible to random seeds, we only use principal component analysis to verify the stock-selection rotation strategy. How to design a suitable mechanism to effectively employ stacked restricted Boltzmann machine and stacked autoencoder for this rotation strategy is a challenge study that we will carry out in the future. Thirdly, how to combine this research with anomalies in behavioral finance is aslo a meaningful work. Last but not the least, how to effectively combine financial econometrics with machine learning to study the stock-selection problem is another topic requiring research.

Acknowledgments

The author would like to thank editors and anonymous reviewers for their constructive comments, and Xiaomin Gong, Jianguo Liu, Qun Chen, Xiaojie Lan, Ziyi Wang for their assistance in writing the article. This work is supported by the Major Program of National Fund of Philosophy and Social Science of China (18ZDA088) and the Fundamental Research Funds for the Central Universities of China (CXJJ-2017-421).

References

Ahmed, P., Lockwood, L.J., Nanda, S., 2002. Multistyle rotation strategies. Journal of Portfolio Management 28, 17–30.

Baser, P., Saini, J.R., 2015. Agent based stock clustering for efficient portfolio management. International Journal of Computer Applications 116, 35–41.

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures, in: Montavon, G., Orr, G., Müller, K.R. (Eds.), Neural networks: Tricks of the trade. 2 ed.. Springer, pp. 437–478.

Brida, J.G., Risso, W.A., 2010. Hierarchical structure of the German stock market. Expert Systems with Applications 37, 3846–3852.

Cai, X., Hu, S., Lin, X., 2012. Feature extraction using restricted boltzmann machine for stock price prediction, in: IEEE International Conference on Computer Science and Automation Engineering, IEEE. pp. 80–83.

Carvalho, C.M., Lopes, H.F., Aguilar, O., 2010. Dynamic stock selection strategies: A structured factor model framework, in: The Ninth Valencia International Meeting, pp. 60–90.

Chong, E., Han, C., Park, F.C., 2017. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications 83, 187–205.

Chong, J., Phillips, G.M., 2012. Low-(economic) volatility investing. The Journal of Wealth Management 15, 75–85.

Cooper, M.J., Gutierrez, R.C., Hameed, A., 2004. Market states and momentum. The Journal of Finance 59, 1345–1365.

Da Costa Jr, N., Cunha, J., Da Silva, S., 2005. Stock selection based on cluster analysis. Economics Bulletin 13, 1–9.

Dary, C.R.M., Shawn, L.K.J., Sabrina, C.H.Y., 2013. Return and risk-return ratio based momentum strategies: A fresh perspective. Journal of Finance and Investment Analysis 2, 1–13.

Ding, C., He, X., Zha, H., Simon, H.D., 2002. Adaptive dimension reduction for clustering high dimensional data, in: IEEE International Conference on Data Mining, IEEE. pp. 147–154.

Dose, C., Cincotti, S., 2005. Clustering of financial time series with application to index and enhanced index tracking portfolio. Physica A: Statistical Mechanics and its Applications 355, 145–151.

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

Fama, E.F., French, K.R., 2017. International tests of a five-factor asset pricing model. Journal of Financial Economics 123, 441–463.

Fama, E.F., French, K.R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Frey, B.J., Delbert, D., 2007. Clustering by passing messages between data points. Science 315, 972–976.

Fulga, C., Dedu, S., Serban, F., 2009. Portfolio optimization with prior stock selection. Economic Computation and Economic Cybernetics Studies and Research 4, 157–171.

Grinblatt, M., Titman, S., Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. The American Economic Review 85, 1088–1105.

Hanna, A.J., 2018. A top-down approach to identifying bull and bear market states. International Review of Financial Analysis 55, 93–110.

Heaton, J.B., Polson, N.G., Witte, J.H., 2017. Deep learning for finance: Deep portfolios. Applied Stochastic Models in Business and Indusry 33, 3–12.

Hinton, G.E., 2012. A practial guide to training restricted boltzmann machines, in: Montavon, G., Orr, G., Müller, K.R. (Eds.), Neural networks: Tricks of the trade. 2 ed.. Springer, pp. 566–619.

Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. Neural Computation 18, 1527–1554.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Hsu, J., Li, F., 2013. Low-volatility investing. Journal of Index Investing 4, 67–72.

Hu, G., Hu, Y., Yang, K., Yu, Z., Sung, F., Zhang, Z., Xie, F., Liu, J., Robertson, N., Hospedales, T.M., et al., 2018. Deep stock representation learning: from candlestick charts to investment decisions, in: International Conference on Acoustics, Speech, and Signal Processing, pp. 2706–2710.

Huang, Z., Heian, J.B., Zhang, T., 2011. Differences of opinion, overconfidence, and the high-volume premium. Journal of Financial Research 34, 1–25.

Iorio, C., Frasso, G., Dambrosio, A., Siciliano, R., 2018. A p-spline based clustering approach for portfolio selection. Expert Systmes with Applications 95, 88–103.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: A review and recent developments. Phil. Trans. R. Soc. A 374, 20150202.

Kirkpatrick, C.D., Dahlquist, J.R., 2010. The basic principle of technical analysis–the trend, in: Technical analysis The complete resource for financial market technicians. 2 ed.. Financial Times Press, pp. 8–19.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, J., Yu, J., 2012. Investor attention, psychological anchors, and stock return predictability. Journal of Financial Economics 104, 401–419.

Li, Y., Wang, S., Tian, Q., Ding, X., 2015. Feature representation for statistical-learning-based object detection: A review. Pattern Recognition 48, 3542–3559.

Lucas, A., Van Dijk, R., Kloek, T., 2002. Stock selection, style rotation, and risk. Journal of Empirical Finance 9, 1–34.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

Nanda, S., Mahanty, B., Tiwari, M., 2010. Clustering Indian stock market data for portfolio management. Expert Systems with Applications 37, 8793–8798.

Parsons, L., Haque, E., Liu, H., 2004. Subspace clustering for high dimensional data: A review. ACM Sigkdd Explorations Newsletter 6, 90–105.

Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications 42, 259–268.

Peachavanish, R., 2016. Stock selection and trading based on cluster analysis of trend and momentum indicators, in: Proceedings of the International MultiConference of Engineers and Computer Scientists, pp. 317–321.

Plyakha, Y., Uppal, R., Vilkov, G., 2014. Equal or value weighting? Implications for asset-pricing tests. Master's thesis. EDHEC Business School.

Ren, F., Lu, Y.N., Li, S.P., Jiang, X.F., Zhong, L.X., Qiu, T., 2017. Dynamic portfolio strategy using clustering approach. Plos One 12, e0169299.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 100, 401–409.

Silva, B., Marques, N.C., 2010. Feature clustering with self-organizing maps and an application to financial time-series for portfolio selection, in: IJCCI (ICFC-ICNC), pp. 301–309.

Steinbach, M., Ertöz, L., Kumar, V., 2004. The challenges of clustering high dimensional data, in: New Directions in Statistical Physics. Springer, pp. 273–309.

Tabak, B.M., Serra, T.R., Cajueiro, D.O., 2010. Topological properties of stock market networks: The case of Brazil. Physica A: Statistical Mechanics and its Applications 389, 3240–3249.

Tajunisha, N., Saravanan, V., 2010. An increased performance of clustering high dimensional data using principal component analysis, in: First International Conference on Integrated Intelligent Computing, pp. 17–21.

Van Der Maaten, L., Postma, E., Van Den Herik, J., 2009. Dimensionality reduction: A comparative review. Journal of Mach Learn Research 10, 66–71.

Verleysen, M., François, D., 2005. The curse of dimensionality in data mining and time series prediction, in: International Work-Conference on Artificial Neural Networks, Springer. pp. 758–770.

Verma, R., Verma, P., 2007. Noise trading and stock market volatility. Journal of Multinational Financial Management 17, 231–243.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Credit Author Statement

Jingti Han: Conceptualization, Format Analysis, Resources, Writing-Review & Editing, Supervision, Project Administration, Funding Acquisition. Zhipeng Ge: Conceptualization, Methodology, Software, Validation, Data Curation, Writing-Original Draft.