Dimensionality Reduction Performance of Sparse PCA Methods

Thanh Do Van(✉)

Faculty of Information Technology, Nguyen Tat Thanh University,
Ho Chi Minh City, Vietnam
dvthanh@ntt.edu.vn

Abstract. Today, a hot topic is building forecasting models on large data sets of economic-financial time-series predictors using dimensionality reduction methods. The forecasting models built based on the dynamic factor model in which the factors extracted by the principal component analysis (PCA for short) method or sparse PCA (SPCA) method are superior to other benchmark models in terms of the forecast accuracy of models. Many pieces of literature have considered that the dimensionality reduction performance of the SPCA method seems to be higher than that of the PCA method. However, there have been no studies comparing the dimensionality reduction performance of those two methods to date.

The purpose of this article is to overcome that inadequacy by experimentally evaluating the dimensionality reduction performance of the two methods mentioned above on ten real-world data sets. Difference from previous beliefs, the experimental results show that the dimensionality reduction performance of the PCA and SPCA methods is competitive.

Keywords: Dimensionality reduction · PCA · Sparse PCA

1 Introduction and Contributions of the Article

The two most prominent approaches used to make forecasts or classifications on large data sets are dataset size reduction and neutron network deep learning. According to [1], neutron network deep learning is only suitable for data sets with a large number of observations, but the number of variables is not too large. The forecasts or classifications on data sets with a large number of predictors or both the number of predictors and the number of observations to be large, first, need to use some techniques to reduce the dimensionality of these data sets, in which the most important is to reduce the number of predictors (or the variable dimension). In this approach, forecasting a target variable on large data sets of time series predictors consists of 2 phases.

The first phase reduces the dimensionality of an original data set by transforming this data set in a high dimensional space to a new data set in a much lower-dimensional space, and phase 2 is to perform a forecast algorithm on the obtained new data set [2]. To reduce the dimensionality of a large data set, one often uses both two techniques of feature selection and feature learning. The feature selection techniques are often used to select highly relevant predictors and nonredundant for forecasting or classifying the

P. Cong Vinh and N. Huu Nhan (Eds.): ICTCC 2021, LNICST 408, pp. 138–148, 2021.
https://doi.org/10.1007/978-3-030-92942-8_12

original data sets [3, 4]. In essence, it is a way to deal with noisy and redundant information in data sets of original predictors. In practical application, the most commonly used feature selection technique is the feature filtering technique using the Pearson correlation coefficient measure or the entropy-based mutual information measure depending on the target variable receiving numerical values or categorical values. The data set of the selected valuable predictors is called the input data set of the predictors. This set is generally still large, and it is necessary to use feature learning techniques to reduce its dimensionality. In socio-economics, the most commonly used feature learning method to reduce the dimensionality of large time-series data sets is the principal component analysis (PCA) method.

The PCA method is an unsupervised learning one. Each principal component is a linear projection of the mean-centered input data set onto an eigenvector of the input data set's covariance matrix [5]. The variance of each principal component is the eigenvalue of a corresponding eigenvector associated with this principal component. Assuming that the principal components of the input data set are sorted in descending order of their respective eigenvalues, then the cumulative variance percentage on the eigenvalue overall of the first k principal components is the percentage of information in the original data set captured by the first k principal components [5]. With a predefined cumulative variance percentage threshold (in the range (70%, 90%)), suppose k is the smallest integer such that the cumulative variance percentage of the first k principal components is greater than this threshold. Then, these k principal components are selected to replace the original predictors in the target variable forecasting exercises according to the original predictors or in the classification exercises of these data sets.

Currently, the building of the target variable forecasting model according to a large set of economic-financial time-series predictors is mainly based on the dynamic factor model where the factors are extracted from the input data set of the predictors using the PCA method [1]. Many empirical studies have shown that the forecast accuracy of models built under such an approach is superior to other benchmark models [6].

Zou et al. [7] argued that since each principal component is a linear combination of all the predictors of the mean-centered input data set and their loadings are often nonzero, it has may have difficulty interpreting the derived principal components. To overcome this limitation, Zou et al. [7] proposed the sparse principal component analysis (SPCA) method, where each principal component is a linear combination of several predictors. According to the preliminary assessment of the authors of this study, the forecast accuracy of forecasting models built based on the dynamic factor model where the factors extracted using the SPCA method can be improved over that the PCA method. Many other researchers also agreed with the above assessment [1, 6, 8]. However, systematic studies have not compared the performance of the dimensionality reduction of the PCA and SPCA methods to confirm the above assessment. The purpose of this article is to experimentally compare the dimensionality reduction performance of the PCA method and the SPCA methods family, including the SPCA, randomized SPCA (RSPCA for short), and Robust SPCA (ROBSPCA) methods. Here the dimensionality reduction performance of a method is measured by the root mean squared forecast error (RMSE for short) of a forecast model built based on the dynamic factor model where the factors are extracted by this dimensionality reduction method.

The article is structured as follows; next, Sect. 2 introduces some preliminaries for the following sections, Sect. 3 presents experimental data sets and methods, Sect. 4 introduces experimental results and evaluations, and the last Sect. 5 presents a few conclusions.

2 Preliminaries

2.1 PCA, SPCA, RSPCA, and ROBSPCA Methods

The PCA Method

Assuming that $\mathbf{X}$ is an input data set. It is a N × m matrix and is expressed as follows $\mathbf{X} = [X_1, X_2, \ldots, X_m]$, here $X_i = (x_{i_1}, x_{i_2}, \ldots, x_{i_N})^{\mathrm{T}} \in \mathbb{R}^{\mathbf{N}}$. **N**, m, and X_i are, respectively, the observation dimension, the variable dimension, and an original variable of $\mathbf{X}$. Denote $\mathrm{Y_h} = (y_{\mathrm{h}_1}, y_{\mathrm{h}_2}, \ldots, y_{h_N})^{\mathrm{T}} \in \mathbb{R}^{\mathbf{N}}$, where h = 1, 2,…, p and $\mathbf{Y} = [\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y_p}]$.

The variable dimension reduction is a mapping $\boldsymbol{R} : \mathbb{R}^{\boldsymbol{m}} \rightarrow \mathbb{R}^{\boldsymbol{p}}$ $(x_{1_j}, x_{2_j}, \ldots, x_{m_j}) \mapsto (y_{1_j}, y_{2_j}, \ldots, y_{p_j})$, s.t. p ≪ m and the data set $[\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y_p}] = \boldsymbol{R}([\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X_m}])$ captures important information of $\mathbf{X}$ as much as possible. As known, when data points $(x_{1_j}, x_{2_j}, \ldots, x_{m_j})$ approximate a hyperplane, the most efficient method for reducing the variable dimension of an $\mathbf{X}$ data set is the PCA [5, 9].

PCA is the unsupervised and typical feature learning method for variable dimension reduction. This method allows transforming high-dimensional data sets into a lower-dimensional subspace while retaining the maximum variance and covariance structure of data sets [5, 10, 11]. The computation of eigenvalues, eigenvectors, and their respective PCs is performed on the covariance matrix of data sets.

Assume that $\mathbf{X} = [X_1, X_2, \ldots, X_m] \in \mathbb{R}^{\mathbf{m}\times\mathbf{N}}$, here $X_i = (x_{i_1}, x_{i_2}, \ldots, x_{i_N})^{\mathrm{T}} \in \mathbb{R}^{\mathbf{N}}$, is a mean-centered matrix, i.e., $\sum_{j=1}^{N} x_{i_j} = 0$, for every i = 1,…, m; $\mathbf{R}$ is the covariance matrix of $\mathbf{X}$; $\mathbf{A}$ is the matrix of the $\mathbf{R}$'s eigenvectors, and $\mathbf{F} = [PC_1, PC_2, \ldots, PC_N] \in \mathbb{R}^{\mathbf{m}\times\mathbf{N}}$ is the matrix of the principal components of $\mathbf{X}$. Then [5]:

$$\mathbf{F} = \mathbf{X}\mathbf{A} \tag{1}$$

The column vectors in $\mathbf{A}$ are loadings vectors. With the A loadings matrix found by the tool of linear algebra as above, the principal components sequentially capture the maximum variability in the input data set. It is also possible to convert finding the $\mathbf{A}$ matrix to a problem of maximum variance. Moreover, the problem of determining the principal components by the PCA method is also found by solving the following problem using variable projection as an optimization strategy [12]:

$$\begin{aligned} &\underset{\mathbf{A}}{\text{minimize}} f(\mathbf{A}) = \tfrac{1}{2}\left\|\mathbf{X} - \mathbf{X}\mathbf{A}\mathbf{A}^{\mathrm{T}}\right\|_{\mathrm{F}}^{2} \\ &\text{subject to } \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I} \end{aligned} \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm of the matrix.

The speed of finding the principal components by finding the solution of the optimal problem (2) is faster than that by the linear algebra approach.

The SPCA and RSPCA Methods

The SPCA method combines sparse regression and standard *PCs* to find a set of sparse loading vectors, i.e., loading vectors with only a few nonzero values [7]. The SPCA method proposed in [12] using variable projection as an optimization strategy is as follows:

$$\begin{aligned} &\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} f(\mathbf{A},\mathbf{B}) = \tfrac{1}{2}\left\|\mathbf{X} - \mathbf{X}\mathbf{B}\mathbf{A}^{\mathrm{T}}\right\|_{\mathrm{F}}^{2} + \psi(\mathbf{B}) \\ &\text{subject to } \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I} \end{aligned} \tag{3}$$

A and **B** are both N × N square matrices where **B** is a sparse weight matrix, and **A** is an orthonormal matrix, ψ denotes a sparsity inducing regularizers such as the LASSO (l_1 norm), or the elastic net. Then the set of factors is formed as **F = XB.**

To speed up the computations, one used matrix approximations to factorize a given matrix into a product of smaller (low-rank) matrices. Erichson et al. [12] formed a low-dimensional sketch of the data, which aims to capture the essential information of the original data set. Then the RSPCA method using variable projection as an optimization strategy can be reformulated as:

$$\begin{aligned} &\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} f(\mathbf{A},\mathbf{B}) = \tfrac{1}{2}\left\|\hat{\mathbf{X}} - \hat{\mathbf{X}}\mathbf{B}\mathbf{A}^{\mathrm{T}}\right\|_{\mathrm{F}}^{2} + \psi(\mathbf{B}) \\ &\text{subject to } \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I} \end{aligned} \tag{4}$$

where $\hat{\mathbf{X}} \in \mathbb{R}^{\mathbf{h}\times\mathbf{N}}$ denotes the sketch of $\mathbf{X} \in \mathbb{R}^{\mathbf{m}\times\mathbf{N}}$**,** and the h dimensions are chosen slightly larger than the target-rank k [12].

The ROSPCA Method

To overcome the challenge that data is grossly corrupted due to measurement errors or other effects in many real-world situations, one often uses the decomposition of a data matrix into its sparse and low-rank components. Erichson et al. [12] proposed another formulation of the ROBSPCA method using variable projection as an optimization strategy. It was also formulated based on separating a data matrix into a low-rank model and a sparse model and introducing prior information like sparsity promoting regularizers. More concretely, the ROBSPCA method using variable projection is as follows:

$$\begin{aligned} &\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} f(\mathbf{A},\mathbf{B}) = \tfrac{1}{2}\left\|\mathbf{X} - \mathbf{X}\mathbf{B}\mathbf{A}^{\mathrm{T}}\right\|_{\mathrm{F}}^{2} + \psi(\mathbf{B}) + \gamma\|\boldsymbol{S}\|_{1} \\ &\text{subject to } \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I} \end{aligned} \tag{5}$$

where matrix **S** captures grossly corrupted outliers in the original data set.

2.2 Dynamic Factor Model

Assume $Y_t, X_{i,t}$ (i = 1, 2,…, N) are stationary time-series [13]. If the frequency of these variables is the same, then the factor bridge equation model and the factor mixed data sampling (MIDAS) model [1, 8, 14, 15] are the autoregressive distributed lag model. This model takes the following form [13]:

$$Y_t = \sum_{k=1}^{S} b_k Y_{t-k} + \sum_{i=1}^{r} \sum_{j=0}^{r_i} \beta_{i_j} X_{i,t-j} + \sum_{j=1}^{p} \sum_{h=0}^{p_j} \gamma_{j_h} F_{j,t-h} + c + u_t \quad (6)$$

Here $X_{i,t}$ is a predictor having a strong influence on the change of the variable Y_t. $F_{j,t}$ is a factor extracted from other predictors; u_t is the model's error and is assumed to be white noise; c, b_k, β_{i_j}, and γ_{j_h} are the estimated parameters; $X_{i,t-j}$ is the variable $X_{i,t}$ lagged j steps. r_i (i = 1,…, r), p_j (j = 1,…,p), and s is the maximum lags of the variables $X_{i,t}$, $F_{j,t}$, and Y_t, respectively. The maximum lag can be determined using the Akaike information criterion (AIC) or the Schwarz information criterion (SIC) [16].

Assume $\hat{Y}_t$ is the fitted variable of the target variable Y_t. *It is* produced by the forecasting model of the target variable. The standard mean forecast error of this model (or RMSE) is determined by [13]:

$$\text{RMSE} = \sqrt{\frac{1}{T} \cdot \sum_{i=1}^{T} \left(Y_i - \hat{Y}_i\right)^2} \quad (7)$$

The smaller the RMSE, the higher the forecast accuracy of the model. RMSE is used to evaluate the performance of a dimensionality reduction method.

3 Experimental Datasets and Method

3.1 Experimental Datasets

Ten data sets in which 03 self-collected data sets and 07 data sets in [17] are used for experiments. Three self-collected data sets are used to forecast the core stock index VN30, consumer price index (CPI), and industrial production output of the Vietnam economy (VIP), and they are named VN30, CPI, and VIP. The remaining seven datasets are named Residential Buildings [18], S&P 500, DJI and Nasdaq [19], Air Quality [20], Energy household volume [21], and SuperConductivity [22]. The VN30, CPI, VIP, and S&P 500 datasets are input data sets, i.e., they do not contain redundant and noisy information for target variable forecasting. The remaining datasets are the original datasets. The S&P 500, DJI, NASDAQ, and Air Quality datasets were supplemented with missing data using the weighted moving average method. The weights depend on each data set. The Residential Building data set is preserved after removing the Zip codes attribute. The S&P 500, DJI, and NASDAQ datasets include observations from November 1, 2010, to October 26, 2017, in their respective original datasets [19], while the dataset The Air Quality data includes observations from noon on March 11, 2004, to noon on April 4, 2005 [20] in the original dataset, where data were collected every hour. The Household Energy dataset includes observations from 5:50 a.m.

on January 11, 2016, to 11:50 a.m. on May 27, 2016, in the original dataset, where data is collected every 10 min [21]. The SuperConductivity data set is the training data set in [22]. Table 1 below shows some statistical characteristics of these data sets. In this table, the number of attributes (No. of Attributes for short) is the number of predictors excluding the target variable.

Table 1. The statistical characteristics of experimental data sets

Data sets	Type of data set	Type of attribute	No. of obser.	No. of attributes	Missing data	The target variable	Freq.
VN30	Time series	Real	366	34	No	VN30 index	Daily
CPI	Time series	Real	72	102	No	CPI index	Monthly
VIP	Time series	Real	60	265	No	Production value of industries	Monthly
Residential building	Multivariate	Real	371	27[a]	No	Sales prices	
S&P500	Time series	Real	1760	52	Yes	S&P 500 index	Daily
DJI	Time series	Real	1760	81	Yes	Dow Jones index	Daily
NASDAQ	Time series	Real	1760	81	Yes	Nasdaq index	Daily
Air quality	Time series	Real	9348	12	Yes	CO of Air	Hourly
Appliances energy	Time series	Real	19704	23	No	The energy use of appliances (wh)	Every 10 min
Super conduct.	Multivariate	Real	21263	81	No	Critical temperature	

[a]Remove the column V1: zip codes

3.2 Experimental Method

In this article, the criterion for selecting the number of extracted factors is their cumulative eigenvalue percentage [23], and forecast models are built based on Eq. (6) using the OLS regression method under ideal conditions. Namely, the maximum lag of all variables in each model is precisely determined using the Akaike information criterion (AIC) [16]. Thus, the maximum lag of the principal components extracted using different variable dimension reduction methods is generally different for each data set. In addition, all variables are tested for unit roots and converted to stationary time series before performing model estimation, and in the final estimated model, all variables are highly statistically significant, at least less than 10%. All conditions for the model's estimate to be the best, linear and unbiased (referred to as BLUE) are guaranteed. That allows improving the forecast accuracy of the estimated model [16]. To conduct the experiments, the article used packages 'Sparsepca' [24], 'Kernlable' [25], and 'Caret' [26] in R.CRAN.

4 Results and Evaluations

With the cumulative eigenvalue percentage threshold of 75% for all the aforementioned variable dimension reduction methods and all the experimental data sets, The variable dimension reduction results include the smallest number of principal components with

the cumulative variance percentage greater than the predefined threshold, a specific percentage of the cumulative variance, the maximum lag of the variables in each forecasting model built based on the factors extracted by methods of PCA, SPCA, RSPCA, ROBSPCA, and RMSE of these models are presented in Table 2.

Table 2. The variable dimension reduction results of the methods

	1. The VN30 dataset				*2. The CPI dataset*			
Methods	PCA	SPCA	RSPCA	ROBSPCA	PCA	SPCA	RSPCA	ROBSPCA
The number of **factors/ PC_s**	14	14	14	15	4	4	4	4
Cumul. Eigen. percentage (%)	75.83	75.8	75.8	76.9	78.72	77.8	77.8	76.7
Maximum lag of variables	5	5	5	5	6	6	6	6
RMSE	0.1895	0.1968	0.1968	0.2054	1.4836	1.0659	1.0673	1.0659
	3. The VIP dataset				*4.The Residence Building*			
Methods	PCA	SPCA	RSPCA	ROBSPCA	PCA	SPCA	RSPCA	ROBSPCA
The number of **factors/ PC_s**	4	4	4	4	1	1	1	1
Cumul. Eigen. percentage (%)	76.19	75.2	75.2	76	99.28	99	99	98.3
Maximum lag of variables	6	6	6	6	10	10	10	10
RMSE	715.96	826.28	1373.57	2642.83	1152.4	1152.5	1152.5	1151.2
	5.The S&P500 data set				*6. The DJI data set*			
Methods	PCA	SPCA	RSPCA	ROBSPCA	PCA	SPCA	RSPCA	ROBSPCA
The number of **factors/ PC_s**	1	1	1	1	1	1	1	1
Cumul. Eigen. percentage (%)	99.96	99.8	99.8	99.8	99.5	99.2	99.2	98.9
Maximum lag of variables	5	5	5	5	5	5	5	5
RMSE	161.415	161.441	161.441	161.441	91.82	309.24	309.24	309.23
	7. The NASDAQ data set				*8. Air Quality data set*			
Methods	PCA	SPCA	RSPCA	ROBSPCA	PCA	SPCA	RSPCA	ROBSPCA
The number of **factors/ PC_s**	1	1	1	1	1	1	1	1
Cumul. Eigen. percentage (%)	99.67	99.4	99.4	99.1	75.96	75.7	75.7	74.8
Maximum lag of variables	5	5	5	5	12	12	12	12
RMSE	365.97	85.47	85.47	85.46	71.459	71.499	71.499	71.427
	9. The Appliances Energy data set				*10. SuperConductivity*			
Methods	PCA	SPCA	RSPCA	ROBSPCA	PCA	SPCA	RSPCA	ROBSPCA
The number of **factors/ PC_s**	3	3	3	3	2	2	2	2
Cumul. Eigen. percentage (%)	79.3	78.7	78.7	78.8	92.3	91.9	91.9	90.5
Maximum lag of variables	6	6	6	6	10	10	10	10
RMSE	101.74	101.76	101.76	101.75	27.314	27.332	27.332	27.319

Extracted from Table 2, Table 3 compares the dimensionality reduction performance of the PCA, SPCA, RSPCA, and ROBCPCA methods performed under the ideal conditions as presented above. To show the dimensionality reduction performance for all ten experimental datasets in the same Figure, the article stretched the data across these 10 data sets by multiplying their data in Table 3 with coefficients 1000, 1, 1, 15, 10, 10, 20, 20 and 100, respectively. Figure 1 contains the column histograms of the stretched data sets.

Table 3. The variable dimension reduction performance of the methods (RMSE)

Methods	D1	D2	D3	D4	D5	D6	D7	D8	D9	D20
PCA	**0.190**	1.484	**715.961**	1152.395	161.415	**91.824**	365.970	71.459	**101.742**	**27.314**
SPCA	0.197	**1.066**	826.276	1152.531	**161.441**	309.241	85.467	71.499	101.764	27.332
RSPCA	0.197	1.067	1373.567	1152.531	**161.441**	309.241	85.467	71.499	101.764	27.332
ROBSPCA	0.205	**1.066**	2642.834	**1151.247**	161.441	309.235	**85.462**	**71.427**	101.747	27.319

In this table, the codes from D1 to D10 are assigned, respectively, to the ten experimental data sets in Table 1 from top to bottom.

Table 3 and Fig. 1 clearly show that in the ten experimental data sets, there are 5/10 cases, where the dimensionality reduction performance of the PCA method is higher than that of the family of the SPCA method, and the remaining 5/10 cases are the opposite, that is, the dimensionality reduction performance of the family of the SPCA method is higher than that of the PCA methods. However, there are many cases where the difference of the dimensionality reduction performance is very insignificant. Namely, there are 5/10 cases (corresponding to Residence Buiding, S&P 500, Air Quality, Appliances Energy, and SuperConductivity datasets) where the dimensionality reduction performance of the PCA, SPCA, RSPCA, and ROBSPCA methods are considered the same.

Table 3 and Fig. 1 also show that in the remaining 5/10 cases, there are 3/10 cases where the dimensionality reduction performance of the PCA method is much higher than that of the family of SPCA methods (corresponding to the VN30, CPI, and NASDAQ datasets). The 2/10 remaining cases where the dimensionality reduction performance of the family of SPCA methods is much higher than that of the PCA method (corresponding to the VIP and DJI data sets).

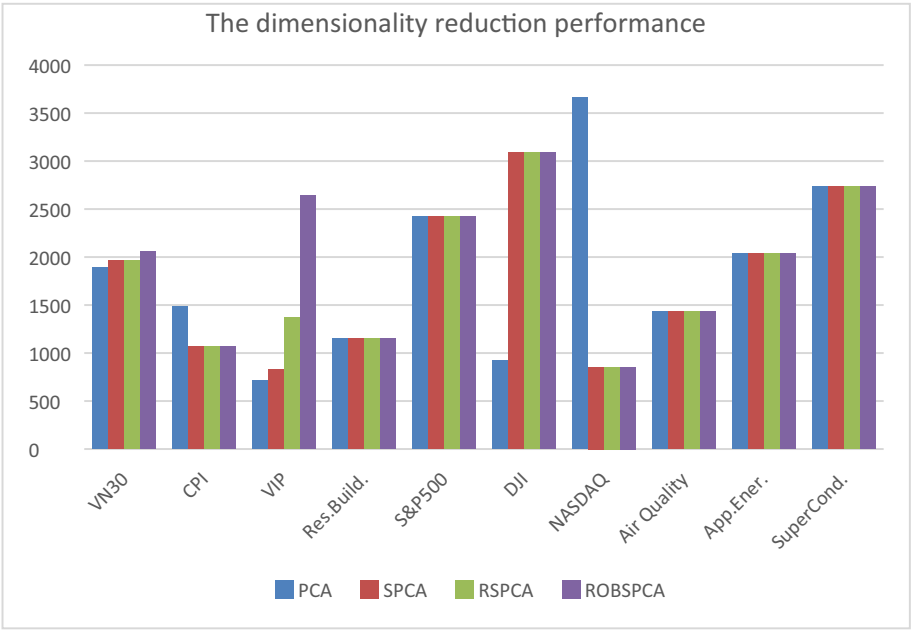

Fig. 1. Compare the dimensionality reduction performance of the methods

In other words, the experiments show that the dimensionality reduction performance of the PCA method and the family of the SPCA method is competitive.

5 Conclusion

This article has experimentally compared the dimensionality reduction performance of the SPCA methods family with the PCA method on ten real-world data sets under ideal conditions. The experimental results show that the dimensionality reduction performance of the SPCA methods family is not higher than the dimensionality reduction performance of the PCA method as long thought by researchers. Their dimensionality reduction performance is competitive.

However, the family of SPCA methods is still worthy of application interest because each principal component found by these methods is only a linear combination of several original predictors. That allows evaluating the impact of some original predictors on the target variable in the context of too many original predictors that can potentially affect the change of the target variable.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Kapetanios, G., Papailias, F., et al.: Big data and macroeconomic nowcasting: methodological review. Economic Statistics Centre of Excellence, National Institute of Economic and Social Research (2018)
2. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024
3. Do Van, T.: Modeling stock price forecasting in the context of high dimensional data. In: Proceedings of the 10th National Science and Technology Conference, FAIR, Da Nang, 17–18 August 2017, pp. 422–433 (2017). (in Vietnamese). https://doi.org/10.15625/vap.2017.00051
4. Do Van, T., Hai, N.M., Hieu, D.D.: Building an unconditional forecast model of Stock Market Indexes using combined leading indicators and principal components: application to the Vietnamese Stock Market. Indian J. Sci. Technol. **11**(2), 1–13. https://doi.org/10.17485/ijst/2018/v11i2/104908
5. Shlens, J.: A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100 (2014)
6. Kim, H.H., Swanson, N.R.: Mining big data using a parsimonious factor, machine learning, variable selection, and shrinkage methods. Int. J. Forecast. **34**(2), 339–354 (2018)
7. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. **15**(2), 265–286 (2006)
8. Chikamatsu, K., Hirakata, N., Kido, Y., Otaka, K., et al.: Nowcasting Japanese GDPs. Bank of Japan (2018)
9. Maaten, L.V.D., Postma, E.: Dimensionality reduction: a comparative review. J. Mach. Learn. Res. **10**, 66–71 (2009)
10. Sorzano, C.O.S., Vargas, J., Montano, A.P.: A survey of dimensionality reduction techniques, pp. 1–35. arXiv preprint arXiv:1403.2877 (2014)
11. Sarveniazi, A.: An actual survey of dimensionality reduction. Am. J. Comput. Math. **04**(02), 55–72 (2014). https://doi.org/10.4236/ajcm.2014.42006
12. Erichson, N.B., Zheng, P., Manohar, K., Brunton, S.L., Kutz, J.N., Aravkin, A.Y.: Sparse principal component analysis via variable projection. SIAM J. Appl. Math. **80**(2), 977–1002 (2020)
13. Koop, G., Quinlivan, R.: Analysis of Economic Data, vol. 2. Wiley, Chichester (2020)
14. Ghysels, E., Santa-Clara, P., Valkanov, R.: The MIDAS touch: mixed data sampling regression models, 34 p. Powered by the California Digital Library University of California (2004). https://escholarship.org/uc/item/9mf223rs
15. Ghysels, E., Santa-Clara, P., Valkanov, R.: Predicting volatility: getting the most out of return data sampled at different frequencies. J. Econom. **131**(1–2), 59–95 (2006)
16. Greene, W.H.: Econometric Analysis, 7th edn. New York University/Prentice-Hall (2012)
17. UCI-Machine Learning Repository. archive.ics.uci.edu/ml/datasets.php. Accessed 5 May 2021
18. Rafiei, M.H., Adeli, H.: A novel machine learning model for estimation of sale prices of real estate units. UCI Machine Learning Repository (2016)
19. Hoseinzade, E., Haratizadeh, S.: CNNpred: CNN-based stock market prediction using a diverse set of variables. UCI Machine Learning Repository (2019)
20. De Vito, S., Massera, E., Piga, M., Martinotto, L., Di Francia, G.: Air quality. UCI Machine Learning Repository (2016)
21. Candanedo, L.M., Feldheim, V., Deramaix, D.: Appliances energy prediction data set. UCI Machine Learning Repository (2017)

22. Hamidieh, K.: Superconductivity data. UCI Machine Learning Repository (2018)
23. Zhang, Y., Li, S., Teng, Y.: Dynamic processes monitoring using recursive kernel principal component analysis. Chem. Eng. Sci. **72**, 78–86 (2012)
24. Erichson, N.B., Zheng, P., Aravkin, S.: Sparsepca: Sparse Principal Component Analysis (SPCA). R package version 0.1.2 (2018). https://CRAN.R-project.org/package=sparsepca
25. Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A.: Kernlab-an S4 package for kernel methods in R. J. Stat. Softw. **11**(9), 1–20 (2004). URL http://www.jstatsoft.org/v11/i09/
26. Kuhn, M., et al.: Building predictive models in R using the caret package. J. Stat. Softw. **28** (5), 1–26 (2008)