Stat Comput (2009) 19: 173–187
DOI 10.1007/s11222-008-9082-y

Principal components analysis of nonstationary time series data

Joseph Ryan G. Lansangan · Erniel B. Barrios

Received: 13 November 2007 / Accepted: 24 June 2008 / Published online: 1 August 2008

Abstract The effect of nonstationarity in time series columns of input data in principal components analysis is examined. Nonstationarity are very common among economic indicators collected over time. They are subsequently summarized into fewer indices for purposes of monitoring. Due to the simultaneous drifting of the nonstationary time series usually caused by the trend, the first component averages all the variables without necessarily reducing dimensionality. Sparse principal components analysis can be used, but attainment of sparsity among the loadings (hence, dimension-reduction is achieved) is influenced by the choice of parameter(s) ($\lambda_{1,i}$). Simulated data with more variables than the number of observations and with different patterns of cross-correlations and autocorrelations were used to illustrate the advantages of sparse principal components analysis over ordinary principal components analysis. Sparse component loadings for nonstationary time series data can be achieved provided that appropriate values of $\lambda_{1,j}$ are used. We provide the range of values of $\lambda_{1,j}$ that will ensure convergence of the sparse principal components algorithm and consequently achieve sparsity of component loadings.

Keywords Principal components analysis · Sparse principal components analysis · Time series · Nonstationarity · Singular value decomposition

J.R.G. Lansangan · E.B. Barrios (✉)
School of Statistics, University of the Philippines Diliman, Magsaysay Ave., Diliman, Quezon City 1101, Philippines
e-mail: ernielb@yahoo.com

J.R.G. Lansangan
e-mail: jrlansangan@yahoo.com

1 Introduction

Principal Components Analysis (PCA) is commonly used in dimension reduction, also a popular tool in index construction. The purpose of PCA is to detect possible structures in the relationships between variables, particularly by reducing the dimensionality of the data using components that capture the information from the different variables (Jollife 2002). PCA finds orthogonal linear combinations of the p original variables, of which a smaller number (usually less than p) explains most of the variability among the original variables. The linear combinations, called the principal components (PCs), are uncorrelated; hence characterization of the PCs, in terms of explained variance is easily done. The technique is commonly used in cross-sectional data for description purposes. Jollife (2002) discussed some issues in PCA of time series data imposing stationarity and using frequency domain analysis. Unlike Factor Analysis that requires a joint distribution of the multivariate observations, PCA can be defined without having to rely on such assumption.

The interpretability of the first few principal components often limits the usefulness of PCA as a descriptive tool. Note that given the input data matrix, PCA is affected primarily by the dependencies between columns and minimally by the dependence within a column (variance). If the input data were observed over time, each column of the data matrix is a time series hence, the temporal dependence in the data is summarized in the diagonal (variances) and off-diagonal (cross-covariances) elements of the variance-covariance matrix. Given that the time series columns of the data matrix are stationary, PCs could still be properly defined since the variance-covariance matrix is not necessarily ill-conditioned. If the time series are nonstationary, the simultaneous drifting can be registered as correlations of the

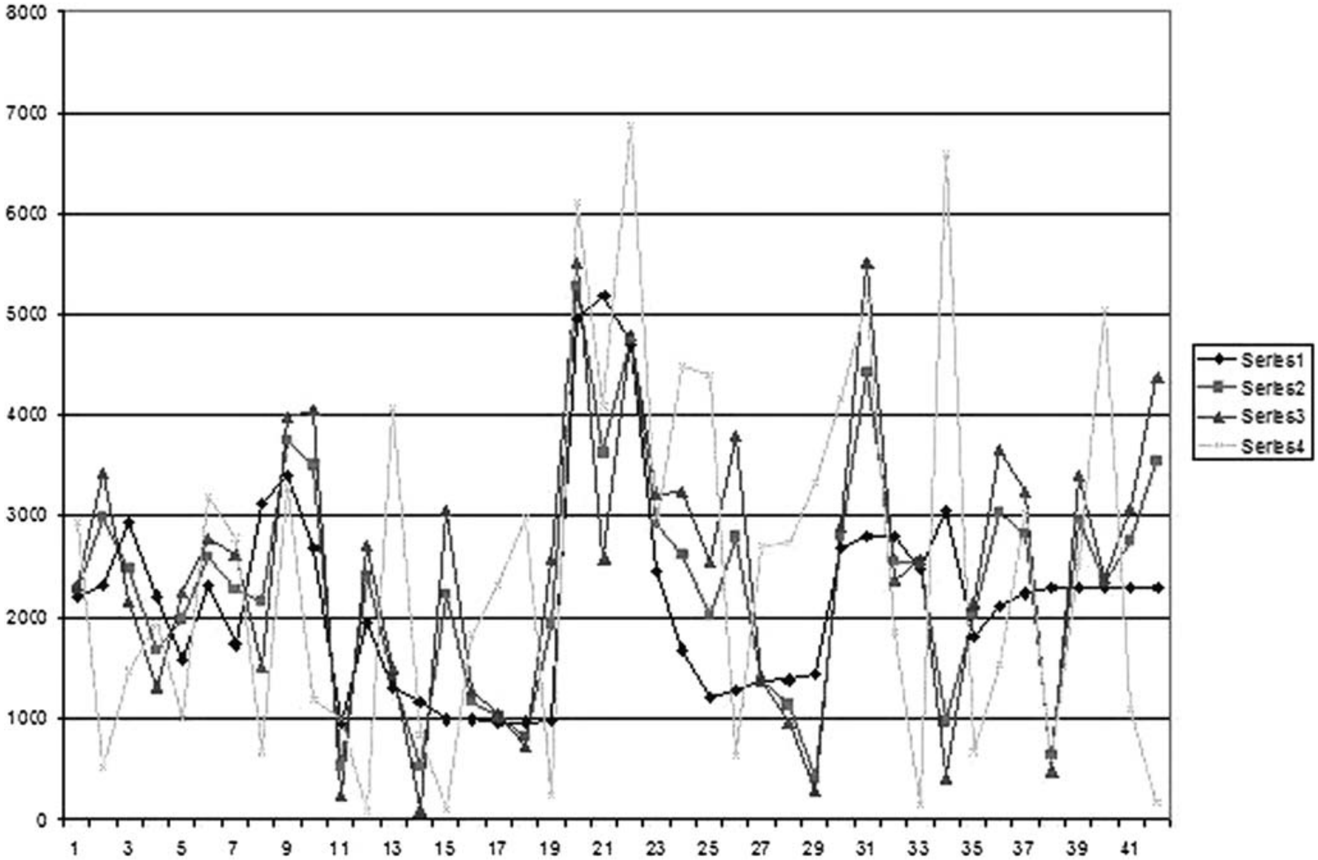

Fig. 1 Time plots of stationary time series

columns of the data matrix. Consider the following illustrations:

The four time series plotted in Fig. 1 are all stationary. The movements of the series over time do not exhibit any pattern that resembles the possibility that they are correlated. In Fig. 2, however, although the time series may not be cointegrated, empirical correlations can be present since the time series drift simultaneously in the same direction, potentially influencing the outcomes of the components. The first component may possibly combine all variables into a single component since variance patterns are clearly similar among all time series. The PCA usually combine together into the same component all variables with similar variability pattern, with similar loadings, indicating equal importance of the variables. Similarity in variance pattern among the variables is usually taken as similarity in the importance of the variables, resulting to the first few components usually "averaging" the variables, i.e., the variables appear to be equally weighted on these components.

Interpretability and sparsity are among the main issues in dimensionality reduction even for cross-sectional data. Jolliffe and Uddin (2000) used both cross-sectional models and pooled time series models to assess and improve new and existing methods of dimension-reduction. In many applications, the number of indicators may exceed the number of observations. PCA in this case serves as a tool in high-dimensional data visualization. Moreover, current techniques for generating components from time-dependent variables assume stationarity of the time series, see for example Jollife (2002), Zuur et al. (2003), Heaton and Solo (2004), and Fernandez-Macho (1997). Gervini and Rousson (2004) proposed some criteria on the assessment of methods of dimension-reduction, while Rousson and Gasser (2004), Vines (2000) and Chipman and Gu (2005) proposed some optimality constraints on the loadings to induce sparsity and hence, interpretability.

Several indicators are used in monitoring to ensure appropriate assessment of the state/status of a phenomenon being monitored. Oftentimes, an intervention is involved pushing the indicators to drift resulting to nonstationarity. But because of the varying patterns among the indicators, a summary is required so that the state of the phenomenon can be reported. This will require index construction, and when principal components analysis is used, this is applied to a set of nonstationary time series data. In this paper, we provide theoretical implications of a nonstationary series, specifically, a drift in mean process, on PCA. To deal with dimension-reduction and interpretability when nonsta-

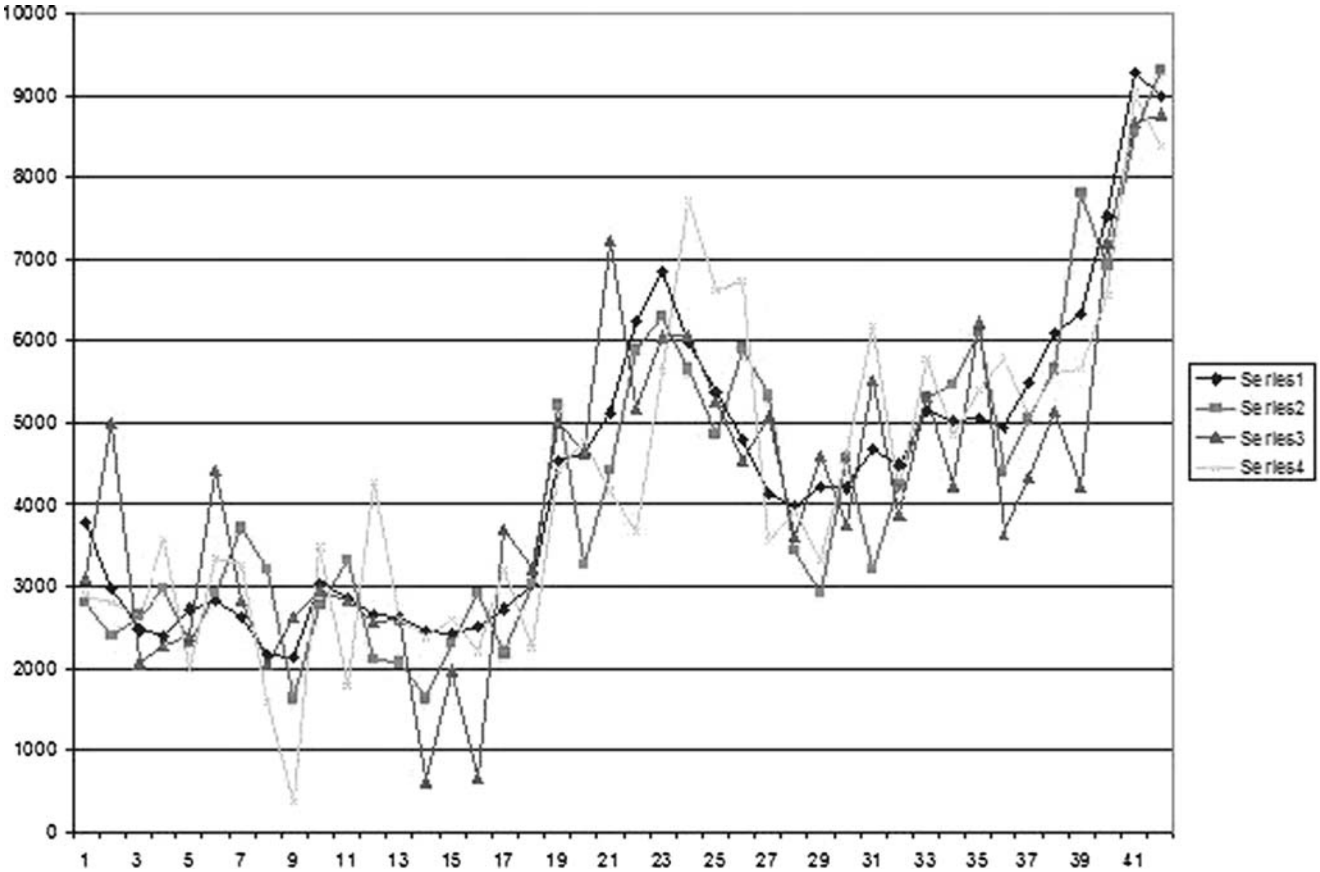

Fig. 2 Time plots of non-stationary time series

tionary data is used, the sparse principal component analysis (SPCA) algorithm proposed by Zou et al. (2006) is applied on simulated nonstationary time series embedded with different cross-correlation and autocorrelation patterns. The next section shows PCA on nonstationary (drift in mean) time series results to a single component. The SPCA formulation and optimization algorithm is briefly discussed in Sect. 3. Applications to simulated data are presented in Sect. 4. Summary of findings and conclusions are in Sect. 5.

2 Effect of nonstationarity in PCA

Consider the first order autoregressive model:

$$y_t = \phi y_{t-1} + \mu_t, \tag{1}$$

where μ_t is a white noise and ϕ is the autoregressive parameter that controls the behavior of the moments of the distribution of y_t. If $|\phi| \geq 1$, the time series is said to be nonstationary (drift in mean). Otherwise, the series is stationary.

Consider the following lemma on the eigenvalues of a matrix:

Lemma *The eigenvalues of a row-unordered $n \times p$ matrix $\underline{X}$ are also the eigenvalues of the row-sequenced matrix $\underline{X}$.*

The lemma follows from the characteristic equation of a matrix. This suggests that the singular value decomposition (SVD) of an input data (where columns are time series) is equivalent to the SVD of the different permutations (with respect to the time points/observations) of the input data. That is, the SVD is invariant to the row ordering of the input data. Hence, the eigenvalues remain the same for any row permutation of the input data even for time series columns.

The following theorem presents the consequences of nonstationarity on ordinary principal components.

Theorem *Let X be a $n \times p$ matrix with columns of a time series following the representation in (1), i.e. $X = [X_1\ X_2\ \ldots\ X_p]$ such that $X_i = \{X_{ti}\}$, $\forall i = 1, 2, \ldots, p$ is the ith time series measured across the time points $t = 1, 2, \ldots, n$. Then, the $p \times p$ diagonal matrix of eigenvalues of X, say D, such that $X = U\ D^{1/2} V'$ for orthonormal matrices U (of dimension $n \times p$) and V (of dimension $p \times p$), is aI_p for some real number a.*

Proof Since X_i is a nonstationary time series, $\forall i = 1, 2, \ldots, p$, then $X_i = \{X_{ti}\} = \{\phi X_{(t-1),i} + \mu_{ti}\}$, where $\mu_{ti} = \mu_i + \varepsilon_{ti}$ with μ_i some constant and $\varepsilon_{ti} \sim N(0, \sigma^2)$, we can write $X_{ti} = \frac{1}{1-\phi B} Y_{ti}$, where $Y_{ti} = \mu_{ti}, \phi$ is the autoregressive parameter (also characterizes stationarity), and B is the backshift operator. X can be written as $\frac{1}{(1-\phi B)} Y$. By SVD, there exists $n \times p$ matrix U and $p \times p$ matrices D and V, where $U^T U = I$, $V^T V = I$, and D is diagonal, such that $X = U D^{1/2} V^T$. D (unique), U and V can be found by first diagonalizing $X^T X$ as $X^T X = V D V^T$ to compute U as $U = X V D^{-1/2}$.

Now, $X^T X = \frac{1}{(1-\phi B)^2} Y^T Y$. But with the existence of $p \times p$ matrices Γ and B, where $B^T B = I$ and Γ is diagonal, $Y^T Y = B \Gamma B^T$. This implies that

$$B \Gamma B^T = V(1 - \phi B)^2 D V^T,$$

$$\Gamma = B^T V(1 - \phi B)^2 D V^T B = S(1 - \phi B)^2 D S^T,$$

where $S = B^T V$,

$$\Gamma^{1/2}\Gamma^{1/2} = S(1 - \phi B)^2 D S^T, \quad \text{since } \Gamma \text{ is diagonal.}$$

Also, $(\Gamma^{1/2})^T \Gamma^{1/2} = S(1 - \phi B)^2 D S^T$, since $\Gamma^{1/2}$ is diagonal and hence symmetric.

$\Gamma^{1/2}$ has eigenvalues $(1 - \phi B) D^{1/2}$ since S is orthonormal and $(1 - \phi B) D^{1/2}$ is diagonal, see Magnus and Neudecker (1999, pp. 18–19, 25).

Since $\Gamma^{1/2} = \{\gamma_{ij}^{1/2}\}$ is itself a diagonal matrix, then $\Gamma^{1/2} = (1 - \phi B) D^{1/2}$. With $D^{1/2} = \{\lambda_{ij}^{1/2}\}$, and that $(1 - \phi B) D^{1/2} = \{(1 - \phi)\lambda_{ij}^{1/2}\}$, this implies that $\gamma_{ii} = \lambda_{ii}(1 - \phi B)^2$. That is, $\Gamma = (1 - \phi B)^2 D$. From Arnold (1981, p. 449), this implies that $\Gamma = a^* I$ for some real number a^*. And hence, $D = aI$, with $a = \frac{a^*}{1-\phi}$. □

The above theorem applied on nonstationary time series gives the following corollaries:

Corollary 1 *Let X be a $n \times p$ matrix of non-stationary (drift in mean) time series, i.e. $X = [X_1\ X_2\ \ldots\ X_p]$ such that $X_i = \{X_{ti}\}$, $\forall i = 1, 2, \ldots, p$, is the ith time series measured across the time points $t = 1, 2, \ldots, n$. Then the eigenvalues of X are undefined.*

Corollary 2 *Let X be a $n \times p$ matrix of non-stationary (drift in mean) time series, i.e. $X = [X_1\ X_2\ \ldots\ X_p]$ such that $X_i = \{X_{ti}\}$ is centered, $\forall i = 1, 2, \ldots, p$, with the ith time series measured across the time points $t = 1, 2, \ldots, n$. Further, let $D = \{\lambda_{ij}\}$ be the $p \times p$ diagonal matrix of eigenvalues of X. Then $\lambda_{11} = \mathrm{tr}(X'X) = p$, $\lambda_{jj} = 0$, $\forall j = 2, 3, \ldots, p$.*

The two corollaries suggest that the PCA of the non-stationary (drift in mean) time series via SVD results to a single component. If the input data consists of non-stationary (drift in mean) time series, a single linear combination of all the time series can solely explain the variability existing within the data. Component loadings for all input variables will be similar if not all equal.

3 Sparse principal component analysis

Zou et al. (2006) proposed an optimization problem that will facilitate the attainment of sparse loadings. They proposed an algorithm which translates PCA in a regression optimization framework and uses the elastic net penalty to derive sparse loadings. Optimization is conducted through regression-type criterion to derive PCs in two stages—first is to perform an ordinary PCA, and second is to find sparse approximations of the first k vector of loadings of the PCs. Let X_i denote the ith row vector of the data matrix $\mathbf{X}$. Let $A_{p\times k} = [\alpha_1, \ldots, \alpha_k]$ and $B_{p\times k} = [\beta_1, \ldots, \beta_k]$.

$$(\hat{A}, \hat{B}) = \arg\min_{A,\beta} \sum_{i=1}^{n} \|X_i - A\beta^T X_i\|^2 + \lambda \sum_{j=1}^{k} \|\beta_j\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \|\beta_j\|_1,$$

subject to $A^T A = I_{k\times k}$.

Whereas the same λ is used for all k components, different $\lambda_{1,j}s$ are allowed for penalizing the loadings of different principal components. The solutions to the optimization problem are called sparse principal components (SPCs).

The proposed modification is reduced to the ordinary PCA when the elastic net penalty is eliminated. But unlike PCA, the algorithm gives components that are correlated and loadings that are not orthogonal. Zou et al. (2006) proposed a computational formula for the total variance explained by the SPCs, which takes into account the correlations of the SPCs. To derive the first k SPCs, values of A are initialized for the minimization of the objective function, i.e., finding solutions to the elastic net problem, then uses singular value decomposition (SVD) to update the loadings. This is then done repetitively until values of β converges. For the case when the number of variables is larger than the number of observations, the algorithm uses soft thresholding over the elastic net optimization.

4 Simulations

The effect of nonstationarity on PCA and SPCA are assessed using simulated data. Different scenarios on non-stationarity

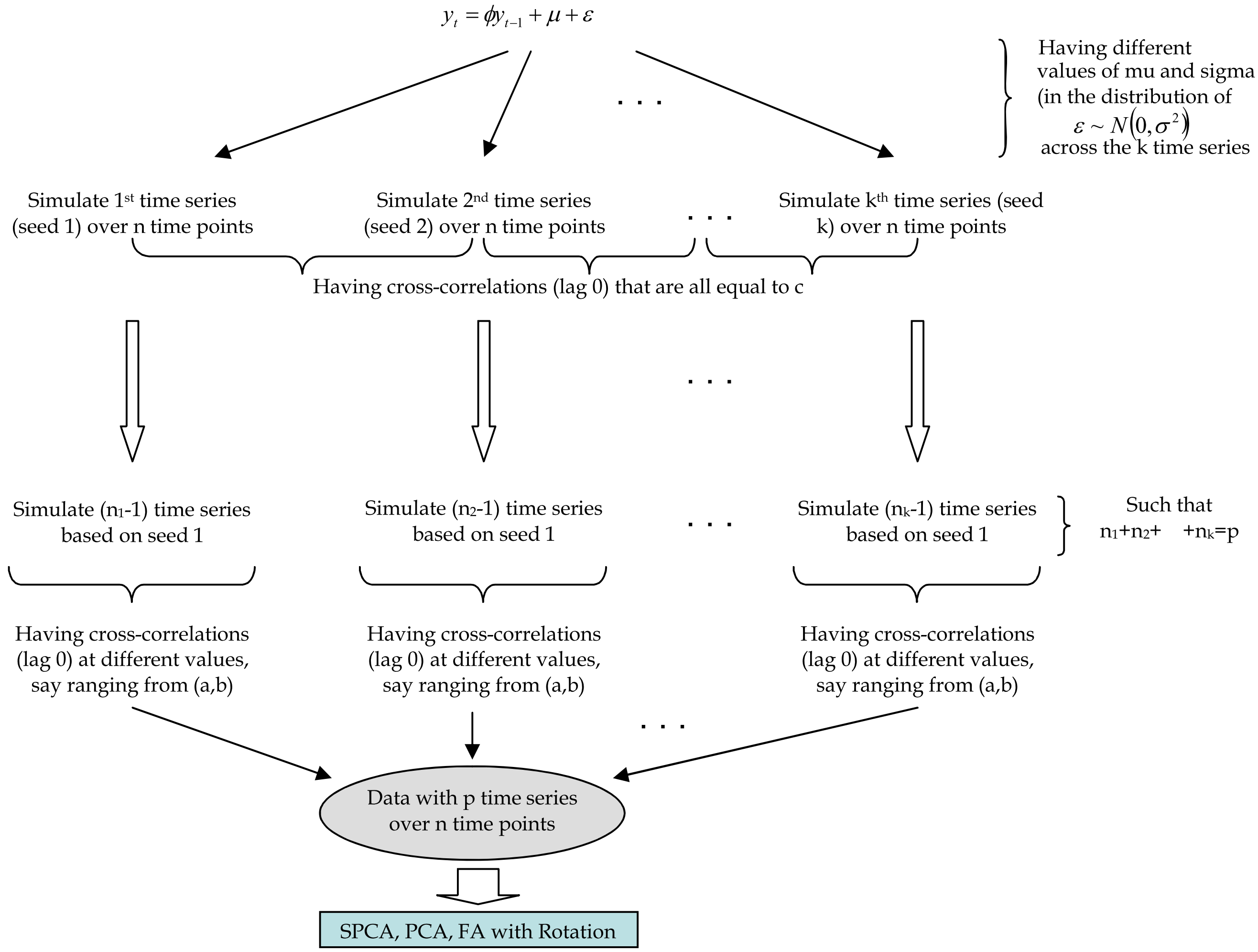

Fig. 3 Simulation process of data with k groups defined with cross-correlations (lag 0) of c

and/or stationarity of the variables and on the level of interdependence among the variables are considered. The simulated data were constructed so that the number of variables exceeds the number of observations. Result of SPCA is then compared to that of PCA.

Given the first order autoregressive model in (1), ϕ were set to 1.3, 1, and 0.7, where $\phi = 0.7$ represents a stationary series, while $\phi = 1.3$ and 1 represent nonstationary series. A set of "similar" time series (i.e., all have the same values of ϕ) were generated. Other groups were generated having different correlations with the first set. The data then consists of the combined groups. Different scenarios are created in terms of the between-group cross-correlations (lag 0). Data with between-group cross-correlations (lag 0) greater than 0.8 (strong), between 0.8 and 0.65 (moderately high), between 0.65 and 0.45 (moderate), between 0.45 and 0.35 (moderately weak), and less than 0.35 (weak), were generated. The simulation procedure for a particular scenario (at given ϕ and between-group cross-correlations, say c) is presented in Fig. 3. The varying interdependencies among the variables between groups are considered to characterize the interaction between nonstationarity and dependencies among the columns of the input data. Note that outcomes of PCA in cross-sectional and stationary time series data are determined primarily by the cross-correlations (lag 0) of the columns of the input data matrix. In the simulation process, the within group cross-correlations (lag 0) were fixed at some range. The reason for grouping the time series is that a set of time series with similar patterns of cross-correlations (lag 0) are expected to dominate the loadings of a principal component. Thus, the number of groupings should coincide with the number of components derived later.

The sparsity of the loadings, interpretability of the components, and the proportion of variance explained by the components from the three procedures were evaluated. The important contribution of the simulation is the identification of cut-offs or intervals for the choice of $\lambda_{1,j}$, $j = 1, 2, \ldots, k$, since some values of $\lambda_{1,j}$ may result to divergence of the algorithm. Contributions to the algorithm of Zou et al. (2006)

Table 1 Summary of simulations

Time series	ϕ	No. of groups	No. of time series per group	No. of time points/obs.	Between-group cross-correlation	Mean of μ_t for ith group	Range of std. dev. of μ_t
Stationary	0.7	3	11	20	High	{500, 550, 600}	(160, 275)
Nonstationary	1.3	3	8	12	High	{500, 100, 200}	(600, 1200)
		2	12	12	Moderately High	{600, 800}	(700, 6500)
		2	12	12	Moderate	{600, 800}	(700, 7150)
		2	12	12	Moderately Weak	{600, 800}	(700, 7150)
		2	12	12	Weak	{600, 800}	(1050, 7150)
	1.0	3	15	30	High	{500, 600, 700}	(400, 960)
		2	19	25	Moderately High	{500, 700}	(560, 2400)
		2	17	20	Moderate	{500, 700}	(540, 1650)
		2	17	20	Moderately Weak	{500, 700}	(900, 1870)
		2	17	20	Weak	{500, 700}	(720, 1980)

Table 2 High correlations for case when ϕ is 0.7

	Minimim cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of SPCs	Range of values for λ (attained sparsity)	No. of zero loadings on SPC1	Range of variance explained by SPCs
Data 1	0.3636	0.9861	0.8246	1	(7.7, 10.2)	1–32	(0.0303, 0.7689)
Data 2	0.4946	0.9682	0.8107	1	(8.1, 10.1)	1–31	(0.0374, 0.7302)
Data 3	0.5213	0.9792	0.8455	1	(8.2, 10.3)	1–31	(0.0377, 0.7965)
Data 4	0.4933	0.9815	0.8262	1	(8.1, 10.2)	1–32	(0.0303, 0.7444)
Data 5	0.4395	0.9752	0.8508	1	(8.2, 10.4)	1–32	(0.0303, 0.7929)

Table 3 High correlations for case when ϕ is 1.3

	Minimim cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of SPCs	Range of values for λ (attained sparsity)	No. of zero loadings on SPC1	Range of variance explained by SPCs
Data 1	0.7123	0.9849	0.8896	1	(8.1, 9.1)	1–22	(0.0638, 0.7338)
Data 2	0.6460	0.9773	0.8738	1	(6.5, 8.9)	1–23	(0.0417, 0.8477)
Data 3	0.7060	0.9790	0.9073	1	(8.3, 9.1)	1–21	(0.0928, 0.8351)
Data 4	0.7110	0.9839	0.8626	1	(7.8, 8.8)	1–23	(0.0417, 0.7578)
Data 5	0.6462	0.9887	0.8861	1	(8.0, 9.0)	1–22	(0.0488, 0.7586)

are on (1) computational tractability for any type of time series, particularly on the number of variables (series) and the number of observations (data points); and (2) sparsity and interpretability (similar variance patterns captured in same component).

The construction of simulated data for the different scenarios—i.e. in terms of the specification of the value of ϕ, number of groups, variables (time series), observations (points), means and variances of the white noise term; is presented in Table 1. Tables 2 to 12 summarize the results for every scenario. The minimum and maximum cross-correlations (Columns 2 and 3) are the lowest and highest observed cross-correlation among all the time series. Column 4 is the proportion of variances explained when the PCs are used. Column 5 gives the number of PCs, and hence SPCs, retained. Column 6 shows the range of possible values of $\lambda_{1,j}$s that give sparse loadings. The second to the last column gives the number of zero loadings in the SPCs. And the last column provides the range of proportion of variance explained by the SPCs.

Table 4 Moderately high correlations for case when ϕ is 1.3

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	0.2048	0.9826	0.8487	2	(5.4, 1.1)–(5.6, 1.7)	SPC1: 2 SPC2: 18	(0.6548, 0.6686)
					(5.6, 2.2)–(6.4, 2.2)	SPC1: 2–6 SPC2: 23	(0.4567, 0.6288)
Data 2	0.1266	0.9609	0.8535	2	(4.5, 1.4)–(4.5, 2.0)	SPC1: 1 SPC2: 14–23	(0.7119, 0.7578)
					(4.9, 1.4)–(4.9, 2.0)	SPC1: 1 SPC2: 14–23	(0.7004, 0.7467)
					(4.5, 2.0)–(7,3, 2.0)	SPC1: 1–14 SPC2: 23	(0.3248, 0.7119)
Data 3	0.0695	0.9777	0.8555	2	(5.1, 2.0)–(5.9, 2.0)	SPC1: 1–2 SPC2: 18–19	(0.5842, 0.6777)
					(5.9, 2.2)–(6.9, 2.2)	SPC1: 2–14 SPC2: 21–23	(0.3104, 0.5602)
					(6.9, 2.4)–(7.1, 2.4)	SPC1: 13–16 SPC2: 23	(0.2622, 0.3124)

Table 5 Moderate correlations for case when ϕ is 1.3

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	−.1355	0.9554	0.7802	2	(3.5, 1.9)–(4.4, 1.9)	SPC1: 1 SPC2: 14	(0.6411, 0.6785)
					(4.4, 2.8)–(6.3, 2.8)	SPC1: 1–18 SPC2: 22–23	(0.1692, 0.5651)
Data 2	−.1588	0.9444	0.7684	2	(3, 1.5)–(3.3, 1.5)	SPC1: 1 SPC2: 10	(0.6850, 0.6941)
					(3.3, 3.0)–(6.7, 3.0)	SPC1: 1–19 SPC2: 23	(0.1730, 0.5949)
Data 3	−.1698	0.9689	0.7821	2	(2.6, 1.3)–(2.6, 1.4)	SPC1: 1 SPC2: 9–10	(0.7289, 0.7353)
					(2.9, 1.4)–(2.9, 3.5)	SPC1: 2 SPC2: 10–23	(0.6232, 0.7274)
					(2.9, 3.5)–(6.6, 3.5)	SPC1: 2–19 SPC2: 23	(0.1677, 0.6232)

Table 6 Moderately weak correlations for case when ϕ is 1.3

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	−.1870	0.9567	0.7532	2	(2.2, 1.7)–(2.2, 3.4)	SPC1: 1 SPC2: 6–22	(0.5582, 0.7140)
					(2.2, 3.4)–(4.5, 3.4)	SPC1: 1–7 SPC2: 22–23	(0.4302, 0.5582)
Data 2	−.0743	0.9439	0.7717	2	(3.0, 1.7)–(3.0, 3.4)	SPC1: 1 SPC2: 9–22	(0.5545, 0.7125)
					(3.0, 3.4)–(5.6, 3.4)	SPC1: 1–13 SPC2: 22	(0.2761, 0.5545)
Data 3	−.4214	0.9233	0.7922	3	(2.1, 1.2, 0.4)–(2.1, 1.2, 1.7)	SPC1: 1 SPC2: 7 SPC3:15–23	(0.7304, 0.7558)
					(2.1, 1.2, 0.4)–(2.1, 1.7, 0.4)	SPC1: 1 SPC2: 7–23 SPC3: 2–15	(0.6831, 0.7558)
					(2.1, 1.2, 1.7)–(2.2, 2.4, 1.7)	SPC1: 1 SPC2: 7–23 SPC3: 23	(0.6065, 0.7304)
					(2.1, 1.2, 0.4)–(5.1, 2.2, 1.6)	SPC1: 1–8 SPC2: 7–18 SPC3: 15–23	(0.4979, 0.7558)
Data 4	−.1870	0.9548	0.7887	3	(3.0, 2.2, 0.9)–(3.0, 2.2, 1.2)	SPC1: 1 SPC2: 11 SPC3: 20–22	(0.6727, 0.6786)
					(3.0, 2.2, 0.9)–(3.1, 2.2, 0.9)	SPC1: 1 SPC2: 11 SPC3: 20	(0.6761, 0.6786)
					(3.1, 2.2, 0.9)–(3.1, 2.2, 1.2)	SPC1: 1 SPC2: 11 SPC3: 20–22	(0.6701, 0.6761)
					(3.1, 2.2, 1.2)–(3.2, 2.4, 1.2)	SPC1: 1 SPC2: 11–13 SPC3: 22	(0.6465, 0.6701)
					(3.2, 2.4, 1.2)–(3.9, 2.4, 1.2)	SPC1: 1–3 SPC2: 13 SPC3: 22	(0.6001, 0.6465)

4.1 Stationary time series

For the stationary case (Table 2), Data 1 to 5 are replications of the simulation process for 20 time points of 33 time series. Factor rotation was not applicable since in all 5 data sets, the PCA resulted to only 1 PC. It is quite expected that PCA will give only a single component because of the high cross-correlations between groups. PCA cannot differenti-

Table 7 Weak correlations for case when ϕ is 1.3

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	−.3993	0.9663	0.8576	2	(1.3, 0.9)–(1.4, 0.9)	SPC1: 8–9 SPC2: 1	(0.7547, 0.7587)
					(4.9, 4.2)–(5.0, 4.2)	SPC1: 10–13 SPC2: 12–16	(0.5461, 0.7041)
Data 2	−.3529	0.9374	0.7773	2	(1.3, 0.9)–(1.4, 1)	SPC1: 10–11 SPC2: 1	(0.6644, 0.6676)
					(5.0, 3.9)–(5.1, 3.9)	SPC1: 12–1 SPC2: 20	(0.4114, 0.4399)

Table 8 High correlations for case when ϕ is 1

	Minimim cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of SPCs	Range of values for λ (attained sparsity)	No. of zero loadings on SPC1	Range of variance explained by SPCs
Data 1	0.7158	0.9912	0.9231	1	(11.2, 12.7)	1–42	(0.0560, 0.8600)
Data 2	0.7240	0.9869	0.9241	1	(11.3, 12.7)	1–43	(0.0436, 0.8733)
Data 3	0.8265	0.9876	0.9301	1	(12.1, 12.7)	1–41	(0.0821, 0.7756)
Data 4	0.8060	0.9884	0.9219	1	(11.7, 12.7)	1–42	(0.0593, 0.8310)
Data 5	0.8014	0.9900	0.9245	1	(11.7, 12.7)	1–40	(0.0929, 0.8276)

Table 9 Moderately high correlations for case when ϕ is 1

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	0.3022	0.9583	0.7703	2	(6.8, 0.4)–(7.5, 0.4)	SPC1: 1–2 SPC2: 32	(0.6727, 0.7130)
					(7.5, 0.9)–(9.3, 0.9)	SPC1: 2–21 SPC2: 37	(0.2891, 0.6635)
Data 2	0.2781	0.9436	0.8026	2	(7.0, 1.6)–(7.7, 1.6)	SPC1: 1–4 SPC2: 33	(0.5817, 0.6603)
					(7.7, 2.0)–(8.7, 2.0)	SPC1: 4–21 SPC2: 37	(0.2877, 0.5667)
Data 3	0.2886	0.9215	0.7744	2	(6.6, 1.6)–(7.3, 1.6)	SPC1: 1–3 SPC2: 33	(0.5418, 0.6284)
					(7.3, 1.8)–(7.8, 1.8)	SPC1: 3–13 SPC2: 34–35	(0.4658, 0.5318)

Table 10 Moderate correlations for case when ϕ is 1

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	0.5019	0.9854	0.8772	2	(5.8, 2.3)– (6.6, 2.3)	SPC1: 1 SPC2: 25	(0.7012, 0.7266)
					(6.6, 3.2)– (7.8, 3.2)	SPC1: 2–12 SPC2: 30–31	(0.5199, 0.6580)
Data 2	0.1330	0.9600	0.8430	2	(5.2, 1.9)– (5.3, 1.9)	SPC1: 1 SPC2: 20	(0.7378, 0.7394)
					(5.3, 2.9)– (7.4, 2.9)	SPC1: 2–8 SPC2: 28–32	(0.5805, 0.6740)
Data 3	0.1787	0.9711	0.8242	2	(5.2, 2.0)– (5.9, 2.0)	SPC1: 1–3 SPC2: 22–23	(0.6865, 0.7011)
					(5.9, 2.8)– (7.1, 2.8)	SPC1: 3–9 SPC2: 27–32	(0.5802, 0.6292)
Data 4	0.1029	0.9287	0.8012	3	(5.8, 1.6, 1.0)– (5.8, 1.6, 1.6)	SPC1: 2 SPC2: 25 SPC3: 29–33	(0.6644, 0.7023)
					(5.8, 1.6, 1.0)– (5.8, 2.4, 1.0)	SPC1: 2 SPC2: 25–33 SPC3: 29	(0.6363, 0.7023)
					(5.8, 1.6, 1.0)– (5.9, 1.6, 1.0)	SPC1: 2 SPC2: 25 SPC3: 29	(0.6935, 0.7023)
					(5.8, 1.6, 1.0)– (6.6, 2.4, 1.0)	SPC1: 2–6 SPC2: 25–33 SPC3: 29	(0.5423, 0.7023)
					(5.8, 1.6, 1.0)– (5.8, 2.4, 1.6)	SPC1: 2 SPC2: 25–29 SPC3: 29–33	(0.6366, 0.7023)
					(5.8, 2.4, 1.6)– (6.7, 2.5, 1.6)	SPC1: 2–7 SPC2: 25–29 SPC3: 29–33	(0.5246, 0.7023)
Data 5	−.0576	0.9300	0.7938	3	(4.1, 1.2, 0.7)– (4.1, 1.2, 1.3)	SPC1: 1 SPC2: 18 SPC3: 23–33	(0.7313, 0.7553)
					(4.1, 1.2, 0.7)– (4.1, 1.5, 0.7)	SPC1: 1 SPC2: 18–22 SPC3: 23	(0.7393, 0.7553)
					(4.1, 1.2, 0.7)– (4.2, 1.2, 0.7)	SPC1: 1 SPC2: 18 SPC3: 23	(0.7543, 0.7553)
					(4.1, 1.2, 0.7)– (4.3, 1.8, 0.7)	SPC1: 1 SPC2: 18–32 SPC3: 15–23	(0.7396, 0.7553)

Table 10 (*Continued*)

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
					(4.1, 1.2, 0.7)– (4.1, 2.9, 1.3)	SPC1: 1 SPC2: 18–33 SPC3: 23–33	(0.6473, 0.7553)
					(4.1, 2.9, 1.3)– (6.2, 2.9, 1.3)	SPC1: 1–5 SPC2: 33 SPC3: 33	(0.5143, 0.6473)
Data 6	0.0078	0.9325	0.8028	3	(4.7, 1.4, 0.8)– (4.7, 1.4, 1.1)	SPC1: 1 SPC2: 19 SPC3: 29–31	(0.7314, 0.7345)
					(4.7, 1.4, 0.8)– (4.7, 1.5, 0.8)	SPC1: 1 SPC2: 19 SPC3: 28–29	(0.7250, 0.7345)
					(5.0, 2.0, 1.0)– (5.0, 2.0, 1.5)	SPC1: 1 SPC2: 28 SPC3: 31–33	(0.6740, 0.6757)
					(5.0, 2.0, 1.0)– (5.0, 2.1, 1.0)	SPC1: 1 SPC2: 28 SPC3: 31	(0.6709, 0.6757)
					(5.0, 2.0, 1.5)– (6.5, 2.0, 1.5)	SPC1: 1–7 SPC2: 28 SPC3: 33	(0.5101, 0.6740)
					(5.0, 2.0, 1.5)– (7.1, 2.4, 1.5)	SPC1: 1–16 SPC2: 28–31 SPC3: 33	(0.3772, 0.6740)

ate the three groups because of high correlations. Although possibly, the correlations within groups are also high, stationary time series may not exhibit within group empirical correlations, and thus resulting to clustering of all the series in a single component. Also, large values of $\lambda_{1,1}$ yield more zero loadings (sparsity), but the sparser the PCs are, the lower the explained proportion of variances is. It can be noted that, for data under such scenario, to obtain sparsity and at least 75% variance explained by the SPCs, $\lambda_{1,1}$ should be at least 7.7.

4.2 Nonstationary time series

Tables 3 to 7 give the summary for the nonstationary time series with $\phi = 1.3$. For the case with between-group cross-correlations that are high (above 0.8), the PCA resulted to only 1 PC, so will the SPCA, and thus factor rotation is not applicable. Again, this is not surprising because of the high cross-correlations. For the other cases, both PCA and SPCA resulted to mostly 2 components, and in some cases, particularly that with between-group cross-correlations less than 0.45, resulted to 3 components. The 3rd SPCs, though, have zero loadings in as much as nearly all of the time series.

Tables 8 to 12 present the results for the nonstationary time series with $\phi = 1$. Note that for the case with high (greater than 0.8) between-group cross-correlations, the time series are summarized to a single component (for PCA and thus, SPCA). For the case with between-group cross-correlations less than 0.8, both PCA and SPCA sug-

Table 11 Moderately weak correlations for case when ϕ is 1

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for $(\lambda_1, \lambda_2, \lambda_3)$ (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	−.0764	0.9548	0.7814	2	(3.6, 2.7)–(4.0, 2.7)	SPC1: 1 SPC2: 12–13	(0.6746, 0.6794)
					(3.6, 2.8)–(4.7, 2.8)	SPC1: 1–2 SPC2: 16–18	(0.6336, 0.6642)
					(3.5, 3.7)–(6.0, 3.7)	SPC1: 1–13 SPC2: 4–13	(0.4326, 0.6789)
Data 2	0.1055	0.9548	0.7886	2	(4.8, 2.7) & (4.8, 2.8)	SPC1: 1 SPC2: 22	0.6295
					(4.8, 2.9)–(5.5, 2.9)	SPC1: 1–3 SPC2: 23	(0.5700, 0.6166)
					(4.6, 3.5)–(5.9, 3.5)	SPC1: 1–18 SPC2: 27–29	(0.4933, 0.5722)
Data 3	−.0859	0.9498	0.7711	2	(4.3, 2.4) & (4.5, 2.4)	SPC1: 1–2 SPC2: 19	(0.6428, 0.6481)
					(4.3, 2.5) & (4.9, 2.5)	SPC1: 1–3 SPC2: 20–21	(0.6209, 0.6368)
					(4.2, 3.6) & (6.5, 3.6)	SPC1: 2–14 SPC2: 29–33	(0.4342, 0.5519)
Data 4	−.1396	0.9030	0.7740	3	(2.4, 2.1, 1.1)–(2.4, 2.1, 2.0)	SPC1: 1 SPC2: 8 SPC3: 33	(0.6854, 0.6856)
					(2.4, 2.1, 1.1)–(2.4, 2.9, 1.1)	SPC1: 1 SPC2: 8–26 SPC3: 33	(0.6007, 0.6856)
					(2.4, 2.1, 1.1)–(3.0, 2.1, 1.1)	SPC1: 1 SPC2: 8–16 SPC3: 33	(0.6792, 0.6856)
					(2.4, 2.1, 1.1)–(5.7, 2.9, 1.1)	SPC1: 1-11 SPC2: 26 SPC3: 33	(0.4693, 0.6856)
					(2.4, 2.1, 1.1)–(2.4, 2.9, 2.0)	SPC1: 1 SPC2: 8–26 SPC3: 33	(0.6038, 0.6856)
					(2.4, 2.9, 2.0)–(5.7, 2.9, 2.0)	SPC1: 1–11 SPC2: 26–27 SPC3: 33	(0.4717, 0.6038)
Data 5	−.0521	0.9218	0.7595	3	(4.1, 2.0, 0.7)–(4.3, 2.0, 0.7)	SPC1: 1 SPC2: 18 SPC3: 31–32	(0.6227, 0.6320)
					(4.1, 2.0, 0.7)–(4.1, 2.2, 0.7)	SPC1: 1 SPC2: 18–21 SPC3: 31	(0.6157, 0.6320)

Table 11 (*Continued*)

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for $(\lambda_1, \lambda_2, \lambda_3)$ (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
					(4.1, 2.0, 0.7)–(4.1, 2.0, 1.2)	SPC1: 1 SPC2: 18–19 SPC3: 31–33	(0.6269, 0.6320)
					(4.1, 2.0, 0.7)–(4.1, 2.8, 1.2)	SPC1: 1 SPC2: 18–30 SPC3: 31–33	(0.5678, 0.6320)
					(4.1, 2.0, 0.7)–(4.3, 2.0, 1.2)	SPC1: 1 SPC2: 18–19 SPC3: 31–33	(0.6182, 0.6320)
					(4.1, 2.0, 0.7)–(4.6, 2.1, 0.7)	SPC1: 1 SPC2: 18–20 SPC3: 31	(0.5985, 0.6320)
Data 5	−.0521	0.9218	0.7595	3	(4.6, 2.1, 0.7)–(4.6, 2.1, 1.2)	SPC1: 3 SPC2: 20 SPC3: 31–33	(0.5948, 0.5985)
					(4.6, 2.1, 1.2)–(6.2, 3.0, 1.2)	SPC1: 3–15 SPC2: 20–30 SPC3: 33	(0.3956, 0.5948)
Data 6	−.2700	0.9161	0.7658	3	(2.9, 1.8, 0.4)–(3.2, 1.8, 0.4)	SPC1: 1 SPC2: 11 SPC3: 24–25	(0.6947, 0.6983)
					(2.9, 1.8, 0.4)–(2.9, 1.8, 1.4)	SPC1: 1 SPC2: 11 SPC3: 24–33	(0.6857, 0.6983)
					(2.9, 1.8, 0.4–(2.9, 3.0, 1.4)	SPC1: 1 SPC2: 11–29 SPC3: 24–33	(0.5923, 0.6983)
					(2.9, 1.8, 0.4)–(3.1, 1.8, 1.4)	SPC1: 1 SPC2: 11 SPC3: 24–33	(0.6833, 0.6983)
					(2.9, 1.8, 0.4)–(3.2, 1.8, 0.4)	SPC1: 1 SPC2: 11 SPC3: 24–25	(0.6947, 0.6983)
					(3.2, 1.8, 0.4)–(6.2, 3.1, 1.4)	SPC1: 1–16 SPC2: 11–28 SPC3: 24–33	(0.4188, 0.6983)

gest two or three components to be retained, but, as in the scenario for which $\phi = 1.3$, the SPCA may zero out almost every loading in the 3rd SPC, depending on the choice of the $\lambda_{1,j}$'s.

4.3 Choice of values for $\lambda_{1,i}$'s in SPCA

Zou et al. (2006) did not suggest how $\lambda_{1,j}$'s can be chosen. It was noted from the simulation exercise that there are

Table 12 Weak correlations for case when ϕ is 1

	Minimum cross-correlation (lag 0) of time series	Maximum cross-correlation (lag 0) of time series	Variance explained by the PCs	No. of PCs/SPCs	Range of values for (λ_1, λ_2) (attained sparsity)	No. of zero loadings on each SPC	Range of variance explained by SPCs
Data 1	−.1517	0.9630	0.8342	2	(1.9, 1.4)–(1.9, 1.5)	SPC1: 17 SPC2: 1	(0.7652, 0.7737)
					(4.1, 2.6)–(4.1, 3.2)	SPC1: 17 SPC2: 13–17	(0.7915, 0.7927)
					(6.3, 4.1)–(6.3, 4.4)	SPC1: 21–24 SPC2: 14–17	(0.5645, 0.6322)
Data 2	−.2203	0.9503	0.8136	2	(2.8, 2.1)–(2.9, 2.1)	SPC1: 17 SPC2: 1	(0.7440, 0.7543)
					(4, 2.6)–(4, 2.9)	SPC1: 17 SPC2: 6–13	(0.7599, 0.7620)
					(4.1, 4.4)–(5.4, 4.4)	SPC1: 2–12 SPC2: 20–26	(0.6031, 0.6464)
Data 3	−.2873	0.9530	0.8025	2	(2.7, 2)–(2.9, 2)	SPC1: 16–17 SPC2: 1	(0.7306, 0.7329)
					(3.5, 2.5)–(3.5, 2.6)	SPC1: 17 SPC2: 6–9	(0.7445, 0.7489)
					(4, 5.3)–(7.1, 5.3)	SPC1: 5–26 SPC2: 2–32	(0.2325, 0.4963)
Data 4	−.1939	0.9611	0.8274	2	(1.5, 1)–(1.6, 1)	SPC1: 15 SPC2: 1	(0.7838, 0.7862)
					(2.7, 1.3)–(2.7, 2.0)	SPC1: 16–17 SPC2: 1–13	(0.7813, 0.8017)
					(6.8, 3.7)–(6.8, 4.8)	SPC1: 33 SPC2: 16–17	(0.4074, 0.4180)
Data 5	−.1252	0.9393	0.7910	2	(4.0, 3.2)–(4.0, 4.3)	SPC1: 1–4 SPC2: 24–27	(0.5912, 0.6321)
					(4.0, 4.3)–(6.7, 4.3)	SPC1: 4–16 SPC2: 27	(0.5077, 0.5912)

values of $\lambda_{1,j}$'s that will not result to convergence of the algorithm, in some cases, it cannot produce sparse component loadings. Based on the simulation results, the recommended intervals for $\lambda_{1,j}$'s are given in Table 13. The intervals are chosen based on the assumptions regarding the existing patterns of the input time series matrix data. The chance of convergence of the algorithm and the attainment of sparse component loadings are higher within these values of $\lambda_{1,j}$'s. Note also that large values of $\lambda_{1,j}$'s give higher number of zero loadings, but at the expense of decreasing the proportion of variance explained by the components.

5 Conclusions

If the input data has, as columns, non-stationary time series, principal components analysis can yield only one or very few components assigning similar loadings to all vari-

Table 13 Intervals for lambdas

No. of sparse components	Intervals for lambdas	Remarks
One	$\lambda_{1,1} = 6.5$–9.1	Use for explosive non-stationary time series with high between-group cross-correlations (lag 0)
	$\lambda_{1,1} = 11.3$–12.7	Use for high (above 0.8) between-group cross-correlations (lag 0)
Two	$\lambda_{1,1} = 4.5$–8.7 $\lambda_{1,2} = 0.4$–2.4	Use for moderately high (0.65–0.8) between-group cross-correlations (lag 0)
	$\lambda_{1,1} = 2.6$–7.8 $\lambda_{1,2} = 1.3$–3.5	Use for time series with moderate (0.45–.65) between-group cross-correlations (lag 0)
	$\lambda_{1,1} = 2.2$–6.5 $\lambda_{1,2} = 1.7$–3.7	Use for time series with moderately weak (0.35–0.45) between-group cross-correlations (lag 0)
	$\lambda_{1,1} = 1.3$–7.1 $\lambda_{1,2} = 0.9$–5.3	Use for time series with weak (below 0.35) between-group cross-correlations (lag 0)
Three	$\lambda_{1,1} = 2.1$–7.1 $\lambda_{1,2} = 1.2$–3.1 $\lambda_{1,3} = 0.4$–1.7	Any non-stationary time series

ables. This lack of sparsity results to difficulty in interpretation of results. This may also yield incorrect information on the state of a phenomenon being monitored through the indicators included in the component (index) lacking sparsity. Sparsity can be attained in constructing principal components of nonstationary time series by imposing constraints on the estimation of the component loadings as proposed by Zou et al. (2006). Dimension-reduction and the search for common patterns among nonstationary time series can be achieved simultaneously. Simulation shows that SPCA can achieve sparsity while consistently recognizing the variance patterns among nonstationary time series. The parameter $\lambda_{1,j}$'s however, should be carefully chosen to ensure convergence of the algorithm.

References

Arnold, S.F.: The Theory of Linear Models and Multivariate Analysis. New York, Wiley (1981)

Chipman, H., Gu, H.: Interpretable dimension reduction. J. Appl. Stat. **32**(9), 969–987 (2005)

Fernandez-Macho, F.: A dynamic factor model for economic time series. Kybernetika **33**(6), 583–606 (1997)

Gervini, D., Rousson, V.: Criteria for evaluating dimension-reducing components for multivariate data. Am. Stat. **58**(1), 72–76 (2004)

Heaton, C., Solo, V.: Identification of causal factor models of stationary time series. Econom. J. **7**(2), 618–627 (2004)

Jolliffe, I.T.: Principal Component Analysis, 2nd ed. Springer, New York (2002)

Jolliffe, I.T., Uddin, M.: The simplified component technique: an alternative to rotated principal components. J. Comput. Graph. Stat. **9**, 689–710 (2000)

Magnus, J.R., Neudecker, H.: Matrix Differential Calculus with Applications in Statistics and Economics. New York, Wiley (1999)

Rousson, V., Gasser, T.: Simple component analysis. Appl. Stat. **53**(4), 539–555 (2004)

Vines, S.: Simple principal components. Appl. Stat. **49**(4), 441–451 (2000)

Zou, H., Hastie, T., Tibshirani, R.: Sparse principal components analysis. J. Comput. Graph. Stat. **15**(2), 265–286 (2006)

Zuur, A.F., Tuck, I.D., Bailey, N.: Dynamic factor analysis to estimate common trends in fisheries time series. Can. J. Fish Aquat. Sci. **60**, 542–552 (2003)