

# Principal Components Analysis of Nonstationary Time Series Data

JOSEPH RYAN G. LANSANGAN and ERNIEL B. BARRIOS

School of Statistics, University of the Philippines Diliman

## Abstract

The effect of nonstationarity in time series columns of input data in principal components analysis is examined. This usually happens when indexing economic indicators for monitoring purposes. The first component averages all the variables without reducing dimensionality. As an alternative, sparse principal components analysis can be used but attainment of sparsity among the loadings is influenced by the choice of a parameter ( $\lambda$ ). Varying cross-correlation and autocorrelation structures were simulated with number of observations exceeding the number of variables. Sparse component loadings even for nonstationary time series columns of the input data can be achieved provided that appropriate value of  $\lambda$  is used. We provide the possible range of values of  $\lambda$  that will ensure convergence of the sparse principal components algorithm and sparsity of component loadings.

*Keywords:* principal components analysis, sparse principal components analysis, time series, non-stationarity, singular value decomposition

*AMS Subject Classification:* 62H25; 91B84, 15A18

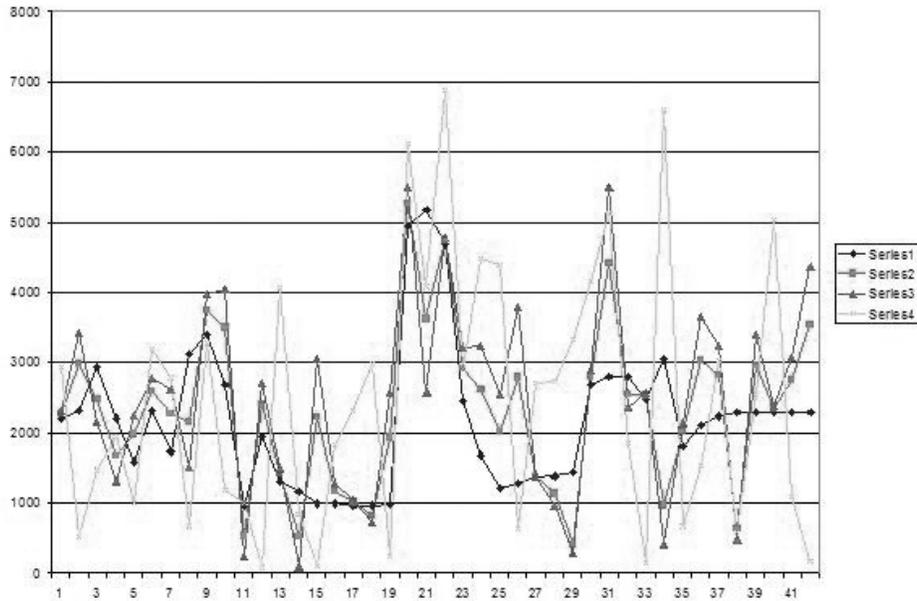
## 1 Introduction

Principal Components Analysis (PCA) is commonly used in dimension reduction and also a popular tool in index construction. The main use 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 in the data brought about by the different variables, (Jolliffe, 2002). PCA finds orthogonal linear combinations of the  $p$  original variables, of which a smaller number (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 quite easily implemented. The technique is commonly employed to cross-sectional data as a descriptive technique. (Jolliffe, 2002) discussed some issues in PCA of time series data following the usual assumption of stationarity and using frequency domain analysis. Unlike Factor Analysis that assumes a joint distribution of the multivariate observations, PCA may be defined without having to impose such assumption.

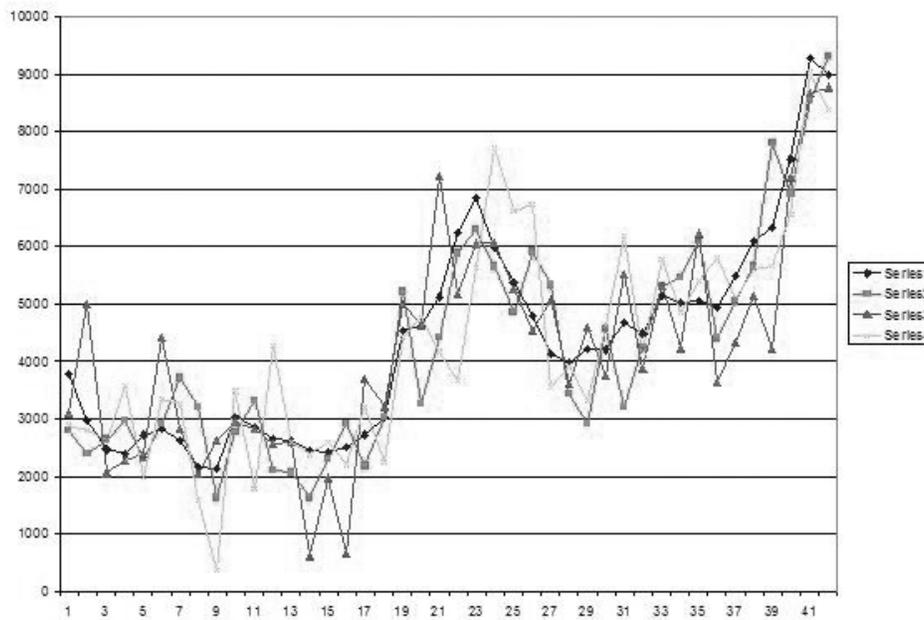
The interpretability of the first few principal components often limits the applications of PCA as a descriptive tool. Note that given the input data matrix, PCA is affected primarily by the dependencies between columns and minimally by the dependence within a column (variance). If the input data were observed over time, each column of the data matrix is time series and 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 ill-conditioning need not manifest in the variance-covariance

matrix. In non-stationary time series however, simultaneous drifting of the series may register as correlations of the columns. Consider the following illustrations:

**Figure 1. Time Plots of Stationary time series**



**Figure 2. Time plots of Non-stationary time series**



In Figure 1 where the four time series are stationary, the patterns may not necessarily indicate correlations. In Figure 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, thus potentially influencing the components. The component may result to possibly combining all variables into a single component since variance patterns are clearly similar. PCA usually combine together into the same component variables with similar variability pattern, with similar loadings indicating the equal importance of the variables. Hence, if the variables have similar variance pattern, this will be taken as similarity in the importance of the variables, resulting to the first few components usually “averaging” the variables, and failure to achieve dimension-reduction.

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 methodologies. In many applications, the number of indicators may exceed the number of observations. Also, current techniques for generating components from time-dependent variables assume stationarity of the time series, see for example, (Jolliffe, 2002), (Zuur et. al., 2003), (Heaton and Solo, 2004), and (Fernandez-Macho, 1997).

In monitoring, several indicators are used to ensure appropriate assessment of the state/status of a phenomenon being monitored. Oftentimes, an intervention is implemented pushing the indicators to drift, resulting to non-stationarity. But because of the varying patterns among the indicators, a summary is needed so that the state of the phenomenon can

be reported. This will require index construction, and principal component analysis needs to be applied to a set of nonstationary time series.

## **2 Some PCA and Related Methods for Time Series**

The use of PCA on time series data is common in the literature. (Jolliffe, 2002) gave several examples applying PCA on time series data. The techniques were dichotomized into time domain and frequency domain. In both cases however, stationarity is assumed. Other applications were discussed in (Lendasse et. al., 2001) which used PCA as a pre-process for forecasting a financial market index.

Dynamic Factor Analysis (DFA) is another dimension reduction technique intended for time series data. DFA can determine whether there exist any underlying common patterns in a set of multiple time series. It also assesses the interactions between the responses and the predictor variables, (Zuur et. al., 2003). In general, instead of modeling individually a set of  $N$  time series for a response variable, a linear combination of  $M$  common trends and the explanatory variables is estimated, with  $M$  taken to be ideally much smaller than  $N$ . Several models can be applied, that is, in terms including a number of common trends, or including explanatory variables and their interactions. (Fernandez-Macho, 1997) used a dynamic factor model that handles nonstationarity via unobserved factors. (Heaton and Solo, 2004) contributed to the identification of a class of dynamic factor model and provided conditions under which the model is identified.

To assess the dimension-reducing effect of a method, (Gervini and Rousson, 2004) proposed some criteria and recommended the one based on the corrected sum of variances (CSV). Sparsity and simplicity was used interchangeably by (Vines, 2000) and proposed a series of “simplicity” preserving linear transformations, resulting to components with integer loadings and usually small values for the first few components. (Rousson and Gasser, 2004) proposed some optimality constraints on the loadings to induce sparsity and hence, interpretability. The resulting components are “sub-optimal” since it explains smaller proportion of variance of the original data, correlated components, but are easy to interpret. (Chipman and Gu, 2005) introduced two classes of constraints: homogeneity and sparsity. Some data illustrates the ability of the procedure to induce sparsity of components in a large set of variables.

The least absolute shrinkage and selection operator (LASSO) was introduced by (Tibshirani, 1996) as a variable selection procedure in regression analysis. The LASSO is a penalized least squares method where the penalty function is an  $L_1$ -norm in the regression coefficients. (Zou and Hastie, 2003) identified some drawbacks of the LASSO, e.g., not well defined unless the bound on the  $L_1$ -norm of the coefficients is smaller than a certain value. They proposed the elastic net as an alternative to the  $L_1$ -norm regularization and pointed out the advantage specially when  $p \gg n$ . (Jolliffe et. al., 2003), developed the simplified component technique-LASSO (SCoTLASS) which is inclusion of the LASSO constraint to the objective function of PCA. Using simulated and actual data, SCoTLASS was preferred over rotated components, and has more interpretable results compared to

ordinary principal components. (Trendafilov and Jolliffe, 2006) developed a globally convergent algorithm for SCoTLASS based on the projected gradient approach. They considered the LASSO constraint as an exterior penalty function. (Tibshirani et. al., 2005) proposed the fused LASSO that included in the penalty function both the  $L_1$ -norm of the coefficients and the successive differences. Sparsity of the coefficients and their differences is attained, especially when  $p \gg n$ .

(Zou et. al., 2006) proposed an optimization problem that will also help obtain sparse loadings. Let  $X_i$  denote the  $i^{\text{th}}$  row vector of the data matrix  $\mathbf{X}$ . Let  $A_{p \times k} = [\alpha_1, \dots, \alpha_k]$  and

$$B_{p \times k} = [\beta_1, \dots, \beta_k]. \quad (\hat{A}, \hat{B}) = \arg \min_{A, B} \sum_{i=1}^n \|X_i - AB^T X_i\|^2 + \lambda \sum_{j=1}^k \|\beta_j\|^2 + \sum_{j=1}^k \lambda_{1,j} \|\beta_j\|_1,$$

subject to  $A^T A = I_{k \times k}$ . Whereas the same  $\lambda$  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 (SPC).

### 3 Effect of Nonstationarity in PCA

Consider the first order autoregressive model:

$$y_t = \phi y_{t-1} + \mu + \varepsilon, \quad (1)$$

where  $\mu$  is some constant,  $\varepsilon$  is a white noise, and  $\phi$  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 non-stationary (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 non-stationarity on ordinary principal components.

*Theorem:* Let  $\underline{X}$  be an  $n \times p$  matrix with columns of a time series following the representation in Equation 1, i.e.  $\underline{X} = [\underline{X}_1 \ \underline{X}_2 \ \cdots \ \underline{X}_p]$  such that  $\underline{X}_i = \{X_{it}\}$ ,  $\forall i = 1, 2, \dots, p$ , is the  $i^{\text{th}}$  time series measured across the time points  $t = 1, 2, \dots, n$ . Then, the  $p \times p$  diagonal matrix of eigenvalues of  $\underline{X}$ , say  $\underline{D}$ , such that  $\underline{X} = \underline{U} \underline{D}^{1/2} \underline{V}^T$  for orthonormal matrices  $\underline{U}$  (of dimension  $n \times p$ ) and  $\underline{V}$  (of dimension  $p \times p$ ), is  $a \underline{I}_p$  for some real number  $a$ .

*Proof:*

Since the  $\underline{X}_i$  is a non-stationary time series,  $\forall i = 1, 2, \dots, p$ , then  $\underline{X}_i = \{X_{it}\} = \{\phi X_{(t-1),i} + \mu_i + \varepsilon_{it}\}$ , where  $\mu_i$  is some constant and  $\varepsilon \sim N(0, \sigma^2)$ , we can

write  $X_{it} = \frac{1}{1 - \phi B} Y_{it}$ , where  $Y_{it} = \mu_i + \varepsilon_{it}$ ,  $\phi$  is the autoregressive parameter (also characterizes stationarity), and  $B$  is the backshift operator.  $\underline{X}$  can be written as

$\frac{1}{(1 - \phi B)} \underline{Y}$ . By SVD, there exists  $n \times p$  matrix  $\underline{U}$  and  $p \times p$  matrices  $\underline{D}$  and  $\underline{V}$ ,

where  $\underline{U}^T \underline{U} = \underline{I}$ ,  $\underline{V}^T \underline{V} = \underline{I}$ , and  $\underline{D}$  is diagonal, such that  $\underline{X} = \underline{U} \underline{D}^{1/2} \underline{V}^T$ .  $\underline{D}$  (unique),

$\underline{U}$  and  $\underline{V}$  can be found by first diagonalizing  $\underline{X}^T \underline{X}$  as  $\underline{X}^T \underline{X} = \underline{V} \underline{D} \underline{V}^T$  to compute  $\underline{U}$

as  $\underline{U} = \underline{X} \underline{V} \underline{D}^{-1/2}$ .

Now,  $\underline{X}^T \underline{X} = \frac{1}{(1 - \phi B)^2} \underline{Y}^T \underline{Y}$ . But with the existence of  $p \times p$  matrices  $\underline{\Gamma}$  and  $\underline{B}$ ,

where  $\underline{B}^T \underline{B} = \underline{I}$  and  $\underline{\Gamma}$  is diagonal,  $\underline{Y}^T \underline{Y} = \underline{B} \underline{\Gamma} \underline{B}^T$ . This implies that

$$\underline{B} \underline{\Gamma} \underline{B}^T = \underline{V} (1 - \phi B)^2 \underline{D} \underline{V}^T$$

$$\underline{\Gamma} = \underline{B}^T \underline{V} (1 - \phi \underline{B})^2 \underline{D} \underline{V}^T \underline{B} = \underline{S} (1 - \phi \underline{B})^2 \underline{D} \underline{S}^T, \text{ where } \underline{S} = \underline{B}^T \underline{V}$$

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

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

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

Since  $\underline{\Gamma}^{1/2} = \{\gamma_{ij}^{1/2}\}$  is itself a diagonal matrix, then  $\underline{\Gamma}^{1/2} = (1 - \phi \underline{B}) \underline{D}^{1/2}$ . With

$\underline{D}^{1/2} = \{\lambda_{ij}^{1/2}\}$ , and that  $(1 - \phi \underline{B}) \underline{D}^{1/2} = \{(1 - \phi) \lambda_{ij}^{1/2}\}$ , this implies that  $\gamma_{ii} = \lambda_{ii} (1 - \phi \underline{B})^2$ .

That is,  $\underline{\Gamma} = (1 - \phi \underline{B})^2 \underline{D}$ . From (Arnold, 1981), this implies that  $\underline{\Gamma} = a^* \underline{I}$  for some real

number  $a^*$ . And hence,  $\underline{D} = a \underline{I}$ , with  $a = \frac{a^*}{1 - \phi}$ .

The above theorem applied on non-stationary time series gives the following corollaries:

*Corollary 1:* Let  $\underline{X}$  be an  $n \times p$  matrix of non-stationary (drift in mean) time series, i.e.

$\underline{X} = [\underline{X}_1 \ \underline{X}_2 \ \dots \ \underline{X}_p]$  such that  $\underline{X}_i = \{X_{it}\}$ ,  $\forall i = 1, 2, \dots, p$ , is the  $i^{\text{th}}$  time series measured

across the time points  $t = 1, 2, \dots, n$ . Then the eigenvalues of  $\underline{X}$  are undefined.

*Corollary 2:* Let  $\underline{X}$  be an  $n \times p$  matrix of non-stationary (drift in mean) time series, i.e.

$\underline{X} = [\underline{X}_1 \ \underline{X}_2 \ \dots \ \underline{X}_p]$  such that  $\underline{X}_i = \{X_{it}\}$  is centered,  $\forall i = 1, 2, \dots, p$ , with the  $i^{\text{th}}$  time

series measured across the time points  $t = 1, 2, \dots, n$ . Further, let  $\underline{D} = \{\lambda_{ij}\}$  be the  $p \times p$  diagonal matrix of eigenvalues of  $\underline{X}$ . Then  $\lambda_{11} = tr(\underline{X}'\underline{X}) = p$ ,  $\lambda_{jj} = 0 \forall j = 2, 3, \dots, p$ .

The corollaries suggest that the PCA of the non-stationary (drift in mean) time series via SVD results to only a single component. That is, 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 input data. Component loadings for all input variables will be similar if not all equal.

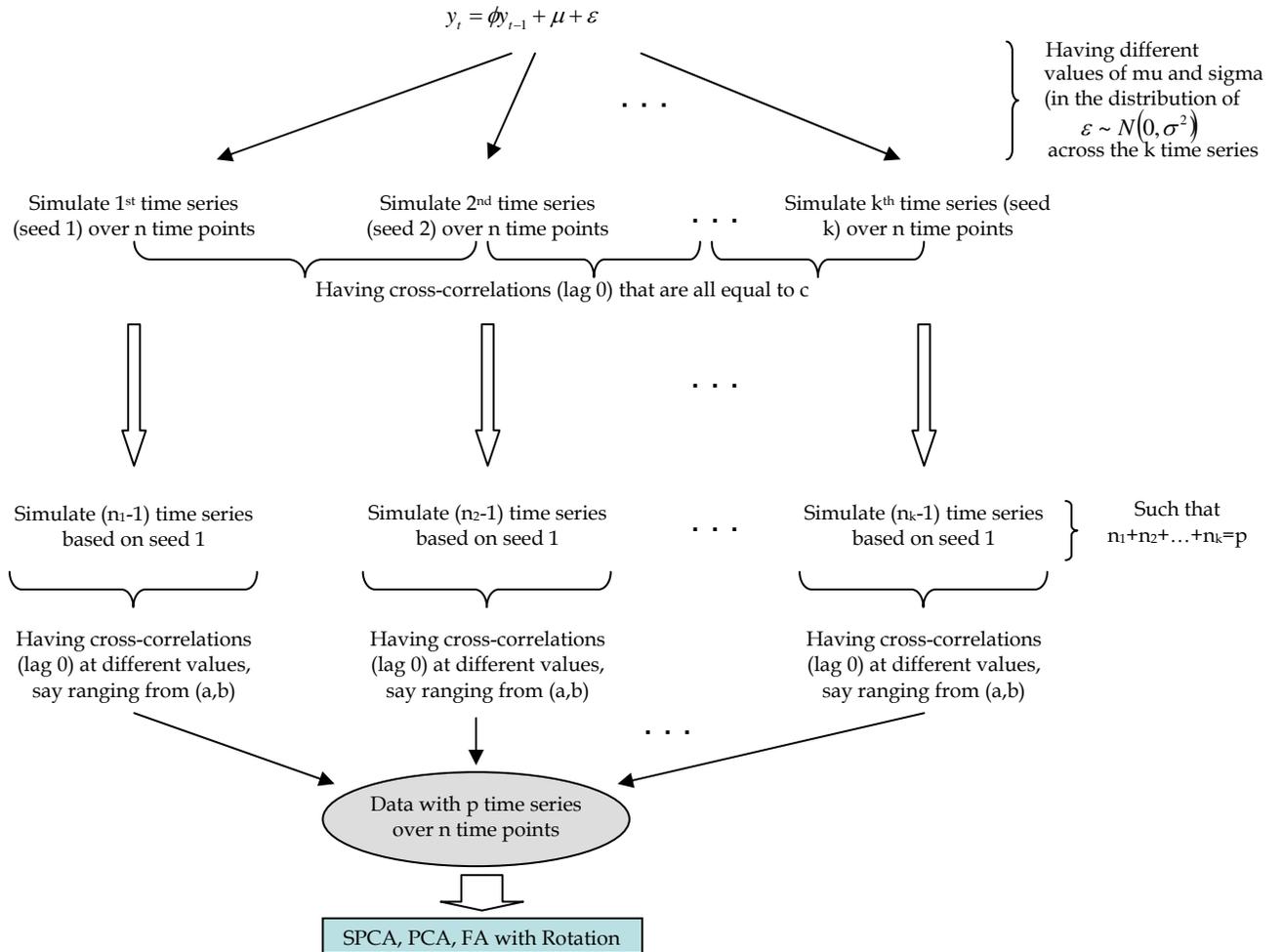
#### **4 Simulations**

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

Given the non-stationary model,  $\phi$  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  $\phi$ ) were generated. Another group 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  $\phi$  and between-group cross-correlations, say  $c$ ) is presented in Figure 3. The varying interdependencies among the variables between groups are considered to characterize the interaction between non-stationarity 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.

**Figure 3. Simulation process of data with  $k$  groups defined with Cross-correlations (lag 0) of  $c$**



The sparsity of the loadings, interpretability of the components, and total predicted variances from the three procedures were evaluated. The important contribution of the simulation is the identification of cut-offs or intervals for the choice of  $\lambda_{1,j}$ ,  $j = 1, 2, \dots, k$ . Some values of  $\lambda_{1,j}$  may result to divergence of the algorithm. Contributions to the algorithm of (Zou et. al., 2006) focuses 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).

Tables 1-11 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 predicted variance 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  $\lambda$  that gives 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 predicted variances of the SPCs.

#### ***4.1 Stationary Time Series: $\phi=0.7$***

PCA generally assigns a component for each group of variables that exhibit similar patterns of variability. The groupings as well as the correlations between and within groups were simulated. Note that  $\phi=0.7$  is a stationary case, hence, only one scenario is simulated, i.e., three groups that are highly correlated.

For each group, the 11 time series are computed from the seed maintaining cross-correlation (lag 0) in the range of 0.40 to 0.99. The series are such that for the  $i^{\text{th}}$  group,  $y_t^{(i)} = \phi y_{t-1}^{(i)} + \mu_i + \varepsilon_i$ , where  $\phi = 0.7$ ,  $\mu_i = (500, 550, 600)$ , and  $\varepsilon_i = N(0, \sigma_i^2)$ , with  $\sigma_i$  ranging from 160 to 275. The results of the PCA and SPCA are summarized in Table 1.

Data 1 to Data 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. This is not surprising, it is quite expected that PCA will give only a single component because of the high cross-correlations between groups. PCA cannot differentiate the 3 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  $\lambda$  yield more zero loadings (sparsity), but the more sparse the PCs are, the lower the predicted variances. It can be noted that, for data under such scenario, to obtain sparsity and at least 75% variance explained by the SPCs,  $\lambda$  should be at least 7.7.

**Table 1. High correlations for case when  $\phi$  is 0.7**

	Minimum 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 $\lambda$ (attained sparsity)	No. of zero loadings on SPC1	Range of Variance Explained by SPCs
<b>Data 1</b>	0.3636	0.9861	0.8246	1	(7.7, 10.2)	1-32	(0.0303, 0.7689)
<b>Data 2</b>	0.4946	0.9682	0.8107	1	(8.1, 10.1)	1-31	(0.0374, 0.7302)
<b>Data 3</b>	0.5213	0.9792	0.8455	1	(8.2, 10.3)	1-31	(0.0377, 0.7965)
<b>Data 4</b>	0.4933	0.9815	0.8262	1	(8.1, 10.2)	1-32	(0.0303, 0.7444)
<b>Data 5</b>	0.4395	0.9752	0.8508	1	(8.2, 10.4)	1-32	(0.0303, 0.7929)

#### **4.2 Nonstationary Time Series: $\phi=1.3$**

For each scenario, analyses were made on different data sets having 12 time points (observations) across 24 time series (variables). For the situation where the between-group cross-correlations (lag 0) are high, 3 groups were considered, having 8 time series in each

group. For the situations where the between-group cross-correlations (lag 0) are moderately high, moderate, moderately weak, or weak, only 2 groups were considered, and each group has 12 time series. The simulation procedure also considered within group cross-correlations. The time series are such that for the  $j^{\text{th}}$  time series in the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = \phi y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where  $\phi = 1.3$ . Adjusting the number of observations to only 12 was considered to maintain the within-group cross-correlations (lag 0) among the time series.

**4.2.1 With 3 Groups That Are Highly Correlated (>0.8)** Note that the time series  $y_t^{(j,i)} = 1.3y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$  are constructed such that there are 3 groups with high between-group cross-correlations (lag 0), i.e., cross-correlations (lag 0) of at least 0.8. And for each group, there are 8 time series with within-group cross-correlations ranging from 0.6 to 0.99. The  $\mu_i$  for the  $i^{\text{th}}$  group is 500, 100, and 200 for  $i = 1, 2, \text{ or } 3$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 600 to 1200.

The results of the PCA and SPCA are summarized in Table 2. Factor rotation was not applicable since in all the data sets, the PCA resulted to only 1 PC. This is again not surprising because the between-group cross-correlations (lag 0) are high.

Table 2. High correlations for case when  $\phi$  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 SPCs	Range of Values for $\lambda$ (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)

#### 4.2.2 With 2 Groups That are Moderately Highly Correlated (Between 0.65 and 0.80)

Note that the time series  $y_t^{(j,i)} = 1.3y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$  are constructed such that there are 2 groups with between-group cross-correlation (lag 0) ranging between 0.65 and 0.80. And for each group, there are 12 time series with within-group cross-correlations ranging from 0.50 to 0.99. The  $\mu_i$  for the  $i^{\text{th}}$  group is 600 and 800 for  $i=1$  or  $2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 700 to 6500. The results of the PCA and SPCA are summarized in Table 3.

Table 3. Moderately high correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	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) -	SPC1: 1	(0.7004, 0.7467)

					(4.9, 2.0)	SPC2: 14-23	
					(4.5, 2.0) - (7.3, 2.0)	SPC1: 1-14 SPC2: 23	(0.3248, 0.7119)
<b>Data 3</b>	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)

#### 4.2.3 With 2 Groups That are Moderately Correlated (Between 0.45 and 0.65) Note

that the time series  $y_t^{(j,i)} = 1.3y_{t-1}^{(j,i)} + \mu_{ij} + \varepsilon_{tij}$  are constructed such that there are 2 groups with between-group cross-correlation (lag 0) ranging between 0.45 and 0.65. And for each group, there are 12 time series with within-group cross-correlations ranging from 0.50 to 0.99. The  $\mu_i$  for the  $i^{\text{th}}$  group is 600 and 800 for  $i=1$  or 2, respectively, and  $\varepsilon_{tij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 700 to 7150. The results of the PCA and SPCA are summarized in Table 4.

**Table 4. Moderate correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	-.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)
<b>Data 2</b>	-.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)
<b>Data 3</b>	-.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) -	SPC1: 2	(0.6232,

					(2.9, 3.5)	SPC2: 10-23	0.7274)
					(2.9, 3.5) - (6.6, 3.5)	SPC1: 2-19 SPC2: 23	(0.1677, 0.6232)

#### 4.2.4 With 2 Groups That are Moderately Weakly Correlated (Between 0.35 and 0.45)

Note that the time series  $y_t^{(j,i)} = 1.3y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$  are constructed such that there are 2 groups with between-group cross-correlation (lag 0) ranging between 0.35 and 0.45. And for each group, there are 12 time series with within-group cross-correlations ranging from 0.40 to 0.99. The  $\mu_i$  for the  $i^{\text{th}}$  group is 600 and 800 for  $i = 1$  or  $2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 700 to 7150. The results of the PCA and SPCA are summarized in Table 5.

Table 5. Moderately weak correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	-0.1870	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)
<b>Data 2</b>	-0.0743	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)
<b>Data 3</b>	-0.4214	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)
<b>Data 4</b>	-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)

#### 4.2.5 With 2 Groups That are Weakly Correlated (Less Than 0.35)

Note that the time series  $y_t^{(j,i)} = 1.3y_{t-1}^{(j,i)} + \mu_{ii} + \varepsilon_{ij}$  are constructed such that there are 2 groups with between-group cross-correlation (lag 0) below 35. And for each group, there are 12 time series with within-group cross-correlations ranging from 0.40 to 0.99. The  $\mu_i$  for the  $i^{\text{th}}$  group is 600 and 800 for  $i=1$  or  $2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 1050 to 7150. The results of the PCA and SPCA are summarized in Table 6.

Table 6. Weak correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	-0.3993	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)
<b>Data 2</b>	-0.3529	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)

#### 4.3 Nonstationary Time Series: $\phi=1$

For the situation where the between-group cross-correlations (lag 0) are high, 3 groups were considered, and having 15 time series in each group. For the situations where the between-group cross-correlations (lag 0) are moderately high, moderate, moderately weak, or weak, only 2 groups were considered. The simulation procedure, as discussed in Section 3.3, also considered within-group cross-correlations. The time series are such that for the  $j^{\text{th}}$  time series in the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = \phi y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where  $\phi = 1.0$ . Adjusting the number of observations for the different scenarios was considered to maintain the within-group cross-correlations (lag 0) among the time series.

**4.3.1 With 3 Groups That are Highly Correlated (>0.8)** For this scenario, analyses were made on 30 time points (observations) across 45 time series (variables). The series are such that for the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where the  $\mu_i$  for the  $i^{\text{th}}$  group is 500,

600, and 700 for  $i = 1, 2, \text{ or } 3$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 400 to 960.

The results of the PCA and SPCA are summarized in Table 7. Factor rotation was not applicable since in any case, the PCA resulted to only 1 PC. This is not surprising because the group correlations are high. Also, just as in the previous cases, large values of  $\lambda$  give higher number of zero loadings, but at the expense of decreasing the predicted variance.

**Table 7. High correlations for case when  $\phi$  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 SPCs	Range of Values for $\lambda$ (attained sparsity)	No. of zero loadings on SPC1	Range of Variance Explained by SPCs
<b>Data 1</b>	0.7158	0.9912	0.9231	1	(11.2, 12.7)	1-42	(0.0560, 0.8600)
<b>Data 2</b>	0.7240	0.9869	0.9241	1	(11.3, 12.7)	1-43	(0.0436, 0.8733)
<b>Data 3</b>	0.8265	0.9876	0.9301	1	(12.1, 12.7)	1-41	(0.0821, 0.7756)
<b>Data 4</b>	0.8060	0.9884	0.9219	1	(11.7, 12.7)	1-42	(0.0593, 0.8310)
<b>Data 5</b>	0.8014	0.9900	0.9245	1	(11.7, 12.7)	1-40	(0.0929, 0.8276)

#### 4.3.2 With 2 Groups that are Moderately Highly Correlated (Between 0.65 and 0.80)

For this scenario, analyses were made on 25 time points (observations) across 38 time series (variables). 2 groups are constructed such that their between-group cross-correlataion (lag 0) is between 0.65 and 0.80. The series are such that for the  $i^{\text{th}}$  group,

$y_t^{(j,i)} = y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where the  $\mu_i$  for the  $i^{\text{th}}$  group is 500 and 700 for  $i = 1, 2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 560 to 2400. The results of the

PCA and SPCA are summarized in Table 8.

Table 8. Moderately high correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	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)
<b>Data 2</b>	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)
<b>Data 3</b>	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)

**4.3.3 With 2 Groups That are Moderately Correlated (Between 0.65 and 0.45)** For this scenario, analyses were made on 20 time points (observations) across 34 time series (variables). That is, the 2 groups that are moderately correlated each have 17 time series.

The series are such that for the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where the  $\mu_i$  for the  $i^{\text{th}}$  group is 500 and 700 for  $i = 1, 2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 540 to 1650. The results of the PCA and SPCA are summarized in Table 9. Results suggest 2 or 3 components to be retained. Though in some cases, there were 3 components generated via ordinary PCA, the SPCA may zero out almost every loading in the 3<sup>rd</sup> SPC, depending on the choice of the lambdas.

Table 9. Moderate correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	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)
<b>Data 2</b>	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)
<b>Data 3</b>	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)
<b>Data 4</b>	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)
<b>Data 5</b>	-.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)
					(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)
<b>Data 6</b>	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)

#### 4.3.4 With 2 Groups that are Moderately Weakly Correlated (Between 0.35 and 0.45)

For this scenario, analyses were made on 20 time points (observations) across 34 time series (variables). The series are such that for the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where the  $\mu_i$  for the  $i^{\text{th}}$  group is 500 and 700 for  $i = 1, 2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 900 to 1870. The results of the PCA and SPCA are summarized in Table 10. Results suggest 2 or 3 components to be retained. Though in some cases, there

were 3 components generated via ordinary PCA, the SPCA zeroes almost every loading in the 3<sup>rd</sup> SPC.

**Table 10. Moderately weak correlations for case when  $\phi$  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 SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	-0.0764	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)
<b>Data 2</b>	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)
<b>Data 3</b>	-0.0859	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)
<b>Data 4</b>	-0.1396	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	(0.4717, 0.6038)

						SPC3: 33	
<b>Data 5</b>	-0.0521	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)
					(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)
<b>Data 5</b>	-0.0521	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)
<b>Data 6</b>	-0.2700	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)

**4.3.5 With 2 Groups that are Weakly Correlated (<0.35)** For this scenario, analyses were made on 20 time points (observations) across 34 time series (variables). The series are such that for the  $i^{\text{th}}$  group,  $y_t^{(j,i)} = y_{t-1}^{(j,i)} + \mu_i + \varepsilon_{ij}$ , where the  $\mu_i$  for the  $i^{\text{th}}$  group is 500 and 700 for  $i = 1, 2$ , respectively, and  $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ , with  $\sigma_{ij}$  ranging from 720 to 1980. The results of the PCA and SPCA are summarized in Table 11. Results suggest 2 principal components to be retained.

**Table 11. Weak correlations for case when  $\phi$  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 $(\lambda_1, \lambda_2)$ (attained sparsity)	No. of zero loadings on SPC1 and SPC2	Range of Variance Explained by SPCs
<b>Data 1</b>	-0.1517	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)
<b>Data 2</b>	-0.2203	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)
<b>Data 3</b>	-0.2873	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)
<b>Data 4</b>	-0.1939	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)
<b>Data 5</b>	-0.1252	0.9393	0.7910	2	(4.0, 3.2) –	SPC1: 1-4	(0.5912, 0.6321)

					(4.0, 4.3)	SPC2: 24-27	
					(4.0, 4.3) - (6.7, 4.3)	SPC1: 4-16 SPC2: 27	(0.5077, 0.5912)

#### 4.4 Choice of Values for $\lambda_{1,i}$ 's in SPCA

(Zou et. al., 2006) did not suggest how  $\lambda_{1,i}$ 's can be chosen. There are values of  $\lambda_{1,i}$ 's however, 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  $\lambda_{1,i}$ 's are given in Table 12. 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  $\lambda_{1,i}$ 's.

**Table 12. Intervals for Lambdas**

No. Of Sparse Components	Intervals for Lambdas	Remarks
<b>One</b>	$\lambda_{1,1} = 6.5 - 9.1$	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)
<b>Two</b>	$\lambda_{1,1} = 4.5 - 8.7$ $\lambda_{1,2} = 0.4 - 2.4$	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)
<b>Three</b>	$\lambda_{1,1} = 2.1 - 7.1$ $\lambda_{1,2} = 1.2 - 3.1$ $\lambda_{1,3} = 0.4 - 1.7$	Any non-stationary time series

## 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 variables. 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 non-stationary time series by imposing constraints on the estimation of the component loadings. 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  $\lambda_{1,i}$ 's however, should be carefully chosen to ensure convergence of the algorithm.

### **Affiliation of Authors:**

#### **Joseph Ryan G. Lansangan**

School of Statistics, University of the Philippines Diliman  
Magsaysay Ave., Diliman, Quezon City 1101  
PHILIPPINES  
Tel/Fax: (63-2) 928 08 81  
**jrlansangan@yahoo.com**

#### **Erniel B. Barrios (Correspondence)**

School of Statistics, University of the Philippines Diliman  
Magsaysay Ave., Diliman, Quezon City 1101  
PHILIPPINES  
Tel/Fax: (63-2) 928 08 81  
**ernielb@yahoo.com**

## References

- Arnold, S. F., 1981, *The Theory of Linear Models and Multivariate Analysis*. (New York: Wiley).
- Chipman, H. and Gu, H., 2005. Interpretable dimension reduction. *J. of Applied Statistics*, **32**(9), 969-987
- Fernandez-Macho, F., 1997, A dynamic factor model for economic time series. *Kybernetika*, **33**(6), 583-606
- Gervini, D., and Rousson, V., 2004, Criteria for evaluating dimension-reducing components for multivariate data. *The American Statistician*, **58**(1), 72-76
- Heaton, C. and V. Solo, 2004, Identification of causal factor models of stationary time series. *The Econometrics Journal*, **7**(2), 618-627
- Jolliffe, I., Trendafilov, N., and Uddin, M., 2003, A modified principal component technique based on the LASSO. *J. of Computational and Graphical Statistics*, **12**(3), 531-547
- Jolliffe, I.T., 2002, *Principal Component Analysis*, 2<sup>nd</sup> ed. (New York: Springer-Verlag).
- Jolliffe, I. T. and M. Uddin, 2000, The simplified component technique: an alternative to rotated principal components. *Journal of Computational and Graphical Statistics*, **9**, 689–710
- Lendasse, A., Lee, J., de Bodt, E, Wertz, V., and Verleysen, M., 2001, Dimension reduction of technical indicators for the prediction of financial time series- application to the BEL20 Market Index. *European Journal of Economic and Social System* **15**(2), 31-48
- Magnus, J.R., and H. Neudecker, 1999, *Matrix Differential Calculus with Applications in Statistics and Economics*. (New York: Wiley).

- Rousson, V. and Gasser, T., 2004, Simple component analysis. *Applied Statistics*, **53**(4), 539-555
- Tibshirani, R., 1996, Regression shrinkage and selection via the LASSO. *J. of the Royal Statistical Soc., Series B*, **58**(1), 267-288
- Tibshirani, R., Saunders, M., Rosset, S., and Zhu, J., 2005, Sparsity and smoothness via the fused LASSO. *J. of the Royal Statist. Soc. B*, **67**(1), 91-108
- Trendafilov, N., and Jolliffe, I., 2006, Projected gradient approach to the numerical solution of the SCoTLASS. *Computational Statistics and Data Analysis*, **50**, 242-253
- Vines, S., 2000, Simple principal components. *Applied Statistics*, **49**(4), 441-451
- Zou, H. and T. Hastie, 2003, Regression and variable selection via the elastic net. *Technical Report*, 2003.
- Zou, H., Hastie, T., and Tibshirani R., 2006, Sparse principal components analysis. *J. of Computational and Graphical Statistics*, **15**(2), 265-286
- Zuur, A.F., Tuck, I.D., and Bailey, N., 2003, Dynamic factor analysis to estimate common trends in fisheries time series. *Can. J. Fish. Aquat. Sci.*, **60**, 542-552