

# Generalized Principal Component Analysis for Moderately Non-Stationary Vector Time Series \*

Fayed Alshammri<sup>a,b</sup> and Jiazhu Pan<sup>a,c</sup>

<sup>a</sup>Department of Mathematics & Statistics, University of Strathclyde, Glasgow G1 1XH, UK

<sup>b</sup>Department of Basic Sciences, Saudi Electronic University, Riyadh, KSA

<sup>c</sup>School of Mathematics & Statistics, Yangtze Normal University, Chongqing 408100, China

## Abstract

This paper extends the principal component analysis (PCA) to moderately non-stationary vector time series. We propose a method that searches for a linear transformation of the original series such that the transformed series is segmented into uncorrelated subseries with lower dimensions. A columns' rearrangement method is proposed to regroup transformed series based on their relationships. We discuss the theoretical properties of the proposed method for fixed and large dimensional cases. Many simulation studies show our approach is suitable for moderately non-stationary data. Illustrations on real data are provided.

**keywords:** Dimension reduction; Eigenanalysis; Moving cross-covariance; Moving cross-correlation; Multivariate time series; Non-stationary data.

**MSC2010 subject classifications:** Primary 62M10; secondary 62H25.

**Declarations of interest:** none.

## 1 Introduction

The collection of multivariate time series data grows as the technology advances in many fields, such as finance, healthcare, industry and social networks. Modelling a multivariate

---

\*Correspondence emails: fayed.alshammri@strath.ac.uk (Fayed Alshammri), Jiazhu Pan: jiazhu.pan@strath.ac.uk (Jiazhu Pan)

---

time series can be a challenge, especially for high dimensional series. Most of the available tools, such as vector autoregressive integrated moving average (VARIMA) models, produce complex and over-parametrized models when applied to such data. Therefore, the use of dimension reduction method is crucial to model high dimensional series and find a simple representation of the data.

There is a significant amount of dimension reduction methods in the literature. A popular approach is the use of principal component analysis (PCA) by projecting the original data into a space with fewer dimensions. Actually, PCA is a commonly used technique to perform dimension reduction for static and independent multivariate data. But the classical PCA will not be able to capture the dynamic dependence of among the components (i.e. variables) of a multivariate time series.

Ku et al. (1995) extended PCA to time series data by including necessary time lags of the original series in the analysis. Their method is called the dynamic principal component analysis (DPCA), and it produces dynamic principal components that are linear combinations of both current and lagged values of the original data. However, DPCA is limited to stationary series, and will not be applicable to nonstationary series. Alshammri and Pan (2019) proposed the moving dynamic principal component analysis (MDPCA), which extended MDPCA to moderately non-stationary time series. Instead of using the cross-covariance matrix, MDPCA applies eigenanalysis on a moving lagged cross-covariance matrix with adjustable window length to extract information from moderately non-stationarity series.

Another related approach was proposed by Brillinger (1981), where the reduction is made based on a reconstruction criterion. This method produces dynamic principal components that are linear combinations of the original series. Peña and Yohai (2016) proposed the generalized dynamic principal component analysis (GDPCA), which is also based on a reconstruction criterion. The dynamic principal components produced by GDPCA could be a nonlinear combination of the original data. The GDPCA can be applied to both stationary and nonstationary series.

Factor models are also widely used tools to reduce the dimension of multivariate time series where the variables of the observed series are considered as linear combinations of some hidden factors that could be interpreted subjectively. See, for example, Peña and Box (1987), Stock and Watson (1988, 2002), Bai and Ng (2002), Forni et al. (2005), Peña and

---

Poncela (2006), Pan and Yao (2008), Lam and Yao (2012) and many others. These models are related to PCA as they are based on eigenanalysis of the covariance matrix of the original series. Zhang et al. (2019) proposed to do co-integration analysis for dimensional reduction of a high dimensional time series with unit root components based a nonnegative definite matrix when there is co-integration relation among components. The co-integration approach is entirely different from PCA. Canonical correlation analysis of Box and Tiao (1977) and scalar component analysis of Tiao and Tsay (1989) are also popular tools to reduce the dimension of a time series.

Of particular interest, a dimension reduction method proposed recently by Chang et al. (2018) called the principal component analysis for time series (TS-PCA). It produces a linear transformation of the original series such that the transformed series is segmented into uncorrelated subseries with lower dimensions. In specific, TS-PCA applies eigenanalysis to a quadratic function of the lagged cross-covariance matrix of the data. The results are uncorrelated subseries (univariate and/or multivariate) that can be analysed separately. However, TS-PCA is limited to second-order stationary series.

In this paper, a new dimension reduction tool will be proposed. The new method we propose is general in the sense that it can be applied to a wide range of both stationary and non-stationary time series data. This method can be considered as an extension of the stationary TS-PCA of Chang et al. (2018) to non-stationary data and therefore will be called the generalized principle component analysis for moderately non-stationary vector time series (GTS-PCA). The results of GTS-PCA are also uncorrelated subseries with lower dimensions that can be analysed individually. We construct a positive definite matrix using a moving window covariance matrix which is different from that for TS-PCA where the classical covariance matrix of the data is used in its calculation. In sections 2 and 3 of this paper we provide a full description of the methodology. Section 2 describes how the GTS-PCA is derived and Section 3 presents a new columns' rearrangement method we propose called the maximum moving cross-correlation method. Theoretical asymptotic properties will be given in Section 4. Performance of the GTS-PCA will be tested in Section 5 on both simulated and real data.

## 2 Methodology

Consider an  $m$ -dimensional time series  $\mathbf{z}_t = (z_{1,t}, z_{2,t}, \dots, z_{m,t})'$  that can be non-stationary.

We are seeking a linear transformation

$$\mathbf{z}_t = \mathbf{A}\mathbf{x}_t \quad (2.1)$$

where  $\mathbf{A}$  is an  $m \times m$  unknown constant matrix and  $\mathbf{x}_t$  is a hidden  $m$ -dimensional time series that is segmented into  $n$  subseries such that  $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{n,t})'$ , where  $1 < n \leq m$ . These  $n$  subseries are uncorrelated with each other. Hence the autocorrelation matrices of  $\mathbf{x}_t$  are block-diagonal matrices with  $n$  being the number of blocks.

Since  $\mathbf{z}_t$  might be non-stationary, the classical cross-covariance function will not be able to truly measure the linear dynamic dependence of  $\mathbf{z}_t$  as it uses a fixed mean vector for the calculation across the whole series. To overcome this problem, we are going to use the moving cross-covariance matrix that was introduced in Alshammri and Pan (2019) as it can measure the linear dynamic dependence of both stationary and non-stationary time series. Before we proceed further, the following definitions are needed.

Define the lag  $l$  cross-covariance matrix of  $\mathbf{z}_i$  as

$$\Gamma_{z,i}(l) = Cov(\mathbf{z}_i, \mathbf{z}_{i-l}) \quad (2.2)$$

where  $l$  is a non-negative integer. Once the series  $\mathbf{z}_t$  is observed, the sample lag  $l$  cross-covariance matrix over window  $\mathbf{z}_i$ , with a pre-specified size of  $2w + 1$ , can be used to estimate  $\Gamma_{z,i}(l)$  as follows

$$\hat{\Gamma}_{z,i}(l) = \frac{1}{2w+1} \sum_{t=i-w}^{i+w} (\mathbf{z}_t - \bar{\mathbf{z}}_i)(\mathbf{z}_{t-l} - \bar{\mathbf{z}}_i)' \quad (2.3)$$

where

$$\bar{\mathbf{z}}_i = \frac{1}{2w+1} \sum_{t=i-w}^{i+w} \mathbf{z}_t,$$

and  $w$  is a positive integer. Then, the lag  $l$  moving cross-covariance matrices of  $\mathbf{z}_t$  and  $\mathbf{x}_t$ , respectively, are defined as

$$\mathbf{M}\Gamma_z(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l, w)+1}^{T-\max(l, w)} \Gamma_{z,i}(l) \quad (2.4)$$

and

$$\mathbf{M}\Gamma_x(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l, w)+1}^{T-\max(l, w)} \Gamma_{x,i}(l),$$

where  $\Gamma_{x,i}(l)$  is defined for  $\mathbf{x}_t$  similarly to that for  $\mathbf{z}_t$ . The sample lag  $l$  moving cross-covariance matrix of  $\mathbf{z}_t$  is calculated as

$$\hat{\mathbf{M}}\Gamma_z(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l, w)+1}^{T-\max(l, w)} \hat{\Gamma}_{z,i}(l) \quad (2.5)$$

while  $\hat{\mathbf{M}}\Gamma_x(l)$  is defined for  $\mathbf{x}_t$  similarly. When  $l = 0$ , we use the notation  $\mathbf{M}\Gamma_z$  and  $\mathbf{M}\Gamma_x$  to refer to the moving covariance matrix of the series  $\mathbf{z}_t$  and  $\mathbf{x}_t$ , respectively. Notice that  $\hat{\mathbf{M}}\Gamma(l)$  in (2.5) extracts its information from the sample lag  $l$  cross-covariance matrices defined over the  $T - 2 \max(l, w)$  windows.

We assume that the moving cross-covariance matrix between any two of the  $n$  uncorrelated subseries  $\mathbf{x}_{j,t}$  and  $\mathbf{x}_{k,t}$  of  $\mathbf{x}_t$  is essentially the zero matrix at any time

$$\mathbf{M}\Gamma(\mathbf{x}_{j,t}, \mathbf{x}_{k,t}) = \mathbf{0} \quad \text{for all } t, s \text{ and } j \neq k. \quad (2.6)$$

For identifiability, we assume without loss of generality that

$$\mathbf{M}\Gamma_z = \mathbf{I}_m \quad \text{and} \quad \mathbf{M}\Gamma_x = \mathbf{I}_m. \quad (2.7)$$

This can be achieved by replacing  $\mathbf{z}_t$  by  $[\hat{\mathbf{M}}\Gamma_z]^{-1/2} \mathbf{z}_t$ . The above assumption will not affect the block structure of  $\mathbf{x}_t$ . Under the above assumptions, the constant matrix  $\mathbf{A}$  is orthonormal since

$$\mathbf{M}\Gamma_z = \mathbf{A} \mathbf{M}\Gamma_x \mathbf{A}' = \mathbf{A} \mathbf{A}' = \mathbf{I}_m \quad (2.8)$$

and therefore, dimension reduction using GTS-PCA will lead to block-diagonal moving autocorrelation matrices of  $\mathbf{x}_t$  with  $n$  blocks. Since  $\mathbf{A}$  is orthonormal, then equation (2.1) can be re-written as

$$\mathbf{x}_t = \mathbf{A}' \mathbf{z}_t. \quad (2.9)$$

For stationary time series, in order to uniquely define  $\mathbf{x}_t$  and  $\mathbf{A}$ , Chang et al. (2018) proposed the use of the sum of a quadratic order of the cross-correlation matrices of  $\mathbf{z}_t$  up to a pre-specified lag  $l_1$ . The result is a positive definite matrix whose eigenvectors' matrix is equivalent to  $\mathbf{A}$  after rearranging its columns. Once  $\mathbf{A}$  is found, the hidden segmentations can be found as the matrix  $\mathbf{x}_t$  can be easily calculated using equation (2.9)

above. This can be extended to account for non-stationary data based on the definitions and the assumptions we have so far as follows. Define the quadratic order moving cross-covariance matrices of  $\mathbf{z}_t$  and  $\mathbf{x}_t$ , respectively, as

$$\mathbf{M}\mathbf{W}_z = \mathbf{I}_m + \sum_{l=1}^{l_1} \mathbf{M}\Gamma_z(l) \mathbf{M}\Gamma_z(l)' \quad (2.10)$$

and

$$\mathbf{M}\mathbf{W}_x = \mathbf{I}_m + \sum_{l=1}^{l_1} \mathbf{M}\Gamma_x(l) \mathbf{M}\Gamma_x(l)'.$$

Thus, we have the following results:

1. The matrices  $\mathbf{M}\mathbf{W}_z$  and  $\mathbf{M}\mathbf{W}_x$  will be the generalized symmetric quadratic order of the cross-covariance matrices used in the stationary TS-PCA and they are also positive-definite matrices.
2. The matrices  $\mathbf{M}\Gamma_x(l)$  and  $\mathbf{M}\mathbf{W}_x$  are block-diagonal  $m \times m$  matrices.
3. It follows from equation (2.1) that

$$\mathbf{M}\mathbf{W}_z = \mathbf{A} \mathbf{M}\mathbf{W}_x \mathbf{A}' \quad (2.11)$$

Once the series  $\mathbf{z}_t$  is observed, the matrix  $\mathbf{M}\mathbf{W}_z$  can be estimated as follows

$$\hat{\mathbf{M}}\mathbf{W}_z = \mathbf{I}_m + \sum_{l=1}^{l_1} \hat{\mathbf{M}}\Gamma_z(l) \hat{\mathbf{M}}\Gamma_z(l)' \quad (2.12)$$

where  $\hat{\mathbf{M}}\Gamma_z(l)$  is defined in (2.5).

Spectral decomposition of the symmetric matrix  $\mathbf{M}\mathbf{W}_x$  yields that

$$\mathbf{M}\mathbf{W}_x = \mathbf{U}_x \Lambda \mathbf{U}_x'$$

where  $\mathbf{U}_x$  is an orthogonal matrix whose columns are the eigenvectors of  $\mathbf{M}\mathbf{W}_x$  and  $\Lambda$  is a diagonal matrix whose diagonal elements are the eigenvalues of  $\mathbf{M}\mathbf{W}_x$ . It follows from equation (2.11) that

$$\mathbf{M}\mathbf{W}_z \mathbf{A} \mathbf{U}_x = \mathbf{A} \mathbf{U}_x \Lambda \quad (2.13)$$

which means that  $\mathbf{U}_z \equiv \mathbf{A} \mathbf{U}_x$  is the orthogonal matrix whose columns are the eigenvectors of  $\mathbf{M}\mathbf{W}_z$  after rearranging its columns. Furthermore, left-multiplying  $\mathbf{z}_t$  by the transposed matrix of  $\mathbf{U}_z$  and noting that  $\mathbf{A}$  is orthonormal we have

$$\mathbf{U}'_z \mathbf{z}_t = \mathbf{U}'_z \mathbf{A} \mathbf{x}_t = \mathbf{U}'_x \mathbf{A}' \mathbf{A} \mathbf{x}_t = \mathbf{U}'_x \mathbf{x}_t. \quad (2.14)$$

The expression in (2.14) exposes the hidden segments, which are the components of the transformed series  $\mathbf{U}'_z \mathbf{z}_t$ . Here,  $\mathbf{U}'_z$  is the transformation matrix used to project the original data into a new space where the transformed data is segmented into  $n$  uncorrelated subseries with lower dimensions that sum up to  $m$ . We assume that different blocks of the quadratic matrix  $\mathbf{M}\mathbf{W}_x$  have different eigenvalues. In the event where different blocks of  $\mathbf{M}\mathbf{W}_x$  have at least one common eigenvalue, then the prospective blocks should be merged together to form one block. Also, following Proposition 1 in Chang et al. (2018),  $\mathbf{U}'_z \mathbf{z}_t$  is equivalent to  $\mathbf{A}' \mathbf{z}_t$  subject to columns' rearrangement.

Since GTS-PCA uses moving cross-covariance matrices in its calculation, then it is important to choose a suitable window size to enhance the results. Obviously, the choice of the window size is case-dependent. In other words, the size depends on the stationarity of the data under study. The more stationary the data, the wider is the window size and the opposite is true. For example, if the data exhibits a strong trend, then a smaller window size is needed. It is also important to choose a window size that is “large enough” to produce reliable results based on consistent estimators. Therefore, when selecting the window size, one should balance between the above issues.

Based on the previous discussion, to reduce the dimension using GTS-PCA and find the hidden segmentation of an observed multivariate time series  $\mathbf{z}_t$ , the following steps can be used:

1. Replace  $\mathbf{z}_t$  in equation (2.1) by  $[\hat{\mathbf{M}}\hat{\Gamma}_z]^{-1/2} \mathbf{z}_t$ .
2. Calculate  $\hat{\mathbf{M}}\hat{\mathbf{W}}_z$  defined in (2.12).
3. Calculate  $\hat{\mathbf{U}}_z$ , the orthogonal matrix that consists of the eigenvectors of  $\hat{\mathbf{M}}\hat{\mathbf{W}}_z$ .
4. Calculate  $\hat{\mathbf{U}}'_z \mathbf{z}_t$ , which is equivalent to  $\mathbf{A}' \mathbf{z}_t$  subject to columns' rearrangement.
5. Find  $\hat{\mathbf{x}}_t = \hat{\mathbf{A}}' \mathbf{z}_t$  by rearranging the columns of  $\hat{\mathbf{U}}'_z \mathbf{z}_t$  such that  $\hat{\mathbf{x}}_t$  is segmented into  $n$  uncorrelated subseries.

### 3 Columns' Rearrangement

The main goal of this section is to find the correct columns' rearrangement of  $\hat{\mathbf{U}}_z$ , which returns  $\hat{\mathbf{A}}$ ; See step 4 and 5 of the GTS-PCA procedure. This can be done by correctly grouping the correlated components of the transformed series  $\hat{\mathbf{U}}_z' \mathbf{z}_t$  in order to find its equivalent  $\hat{\mathbf{A}}' \mathbf{z}_t$ . For ease of notation, denote  $\hat{\mathbf{f}}_t = \hat{\mathbf{U}}_z' \mathbf{z}_t$ . The following strategy can be used to reach our goal. Start with each component of  $\hat{\mathbf{f}}_t$  as a separate group. Then merge any two correlated groups into one group. Repeat the last step until no further groups are correlated. Examining for significant correlation between the variables of a multivariate time series can be done by visualizing tools such as the moving cross-correlation plots introduced in Alshammri and Pan (2019). Other methods in literature are based on ratio criteria instead of using a pairwise multiple hypothesis testing. These ratio tests are favourable when the dimension of the series is large because they use less calculation to find correlated components compared with the other methods. Methods such as co-integration searches for stationary linear combinations of non-stationary series. However, co-integration is concerned with the long-run relationships between non-stationary variables; See Engle and Granger (1987), Johansen (1995) and Zhang et al. (2019). Chang et al. (2018) developed the maximum cross-correlation method that can capture the dynamic dependence of stationary series. But this method is not suitable for non-stationary series because it uses the classical correlation function in its calculations.

For GTS-PCA, we are going to use the moving cross-correlation function of Alshammri and Pan (2019) to find correlated components. Additionally, a new ratio based method will be developed to detect the correlated components of series with high dimensions.

#### 3.1 The Moving Cross-Correlation Function

The moving cross-correlation function is used to examine the dynamic dependence between the components (i.e. variables) of both stationary and non-stationary vector time series. Here, this function will be mainly used to investigate the relationships between the  $m$  components of the transformed series  $\hat{\mathbf{f}}_t$  in order to correctly group the correlated components and find the hidden segmentations. Define the lag  $l$  moving cross-correlation matrix of  $\mathbf{z}_t$  as

$$\mathbf{M}\rho_z(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l, w)+1}^{T-\max(l, w)} \mathbf{S}_i^{-1} \Gamma_{z,i}(l) \mathbf{S}_{i-l}^{-1} \quad (3.1)$$

where  $\Gamma_{z,i}(l)$  is defined in (2.2) and  $\mathbf{S}_i$  is the diagonal matrix of the standard deviations of  $\mathbf{z}_i$ . The  $(j, j)^{th}$  element of  $\mathbf{S}_i$  is the square root of the  $(j, j)^{th}$  element of  $\Gamma_{z,i}(0)$  defined over  $\mathbf{z}_i$ . Denote the  $(j, k)^{th}$  element of  $\mathbf{M}\rho_z(l)$  by  $\mathbf{M}\rho_{z,jk}(l)$ , which is the lag  $l$  moving correlation between the components  $z_{j,t}$  and  $z_{k,t-l}$ . Moreover, the sample lag  $l$  moving cross-correlation matrix can be calculated to estimate  $\mathbf{M}\rho_z(l)$  as follows

$$\hat{\mathbf{M}}\rho_z(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l, w)+1}^{T-\max(l, w)} \hat{\mathbf{S}}_i^{-1} \hat{\Gamma}_{z,i}(l) \hat{\mathbf{S}}_{i-l}^{-1} \quad (3.2)$$

where  $\hat{\Gamma}_{z,i}(l)$  is defined in (2.3), and the  $(j, j)^{th}$  element of  $\hat{\mathbf{S}}_i$  is the square root of the  $(j, j)^{th}$  element of  $\hat{\Gamma}_{z,i}(0)$ .

Notice that the moving cross-correlation matrix has the following property

$$\mathbf{M}\rho_z(l) = [\mathbf{M}\rho_z(-l)]'. \quad (3.3)$$

Therefore, in order to calculate and plot the moving cross-correlation between the components of the series  $\mathbf{z}_t$  in different time lags, it is enough to calculate  $\mathbf{M}\rho_z(l)$  for positive lags (i.e.  $l \geq 0$ ), then use the property in (3.3) to calculate  $\mathbf{M}\rho_z(l)$  for negative lags (i.e.  $l < 0$ ). The function in (3.1) will be used to evaluate the relationship between the components produced by GTS-PCA. For further information and examples about the usage of the moving cross-correlation function and its plots, the reader is referred to Alshammri and Pan (2019).

For the case where the time series has a large number of variables, the moving cross-correlation plots will not be practical as there will be  $m(m - 1)$  individual plots to be examined for a series with dimension  $m$ . Hence, the use of multiple hypothesis testing or a ratio-based method is recommended. Therefore, we propose a new method which is called the maximum moving cross-correlation method.

### 3.2 The Maximum Moving Cross-Correlation Method

The maximum moving cross-correlation method is a ratio-based test that will be used to test for significant correlations between the variables of both stationary and non-stationary

series. In general, for an observed series  $\mathbf{z}_t$ , we consider its  $j^{\text{th}}$  and  $k^{\text{th}}$  variables to be correlated if  $\hat{\mathbf{M}}\rho_{z,jk}(l)$ , the  $(j, k)^{\text{th}}$  element of  $\hat{\mathbf{M}}\rho_z(l)$ , is significantly different from zero over different time lags. In other words, the two components  $\mathbf{z}_{j,t}$  and  $\mathbf{z}_{k,t}$  are correlated if we reject the hypothesis

$$H_o : \mathbf{M}\rho_{z,jk}(l) = 0 \quad \text{for } l = 0, \pm 1, \pm 2, \dots, \pm p \quad (3.4)$$

against

$$H_a : \mathbf{M}\rho_{z,jk}(l) \neq 0 \quad \text{for some } -p \leq l \leq p$$

where  $p$  is a prescribed positive integer.

The maximum moving cross-correlation statistics between components  $\mathbf{z}_{j,t}$  and  $\mathbf{z}_{k,t}$  is calculated as

$$\hat{\mathbf{L}}_d(j, k) = \max_{-p \leq l \leq p} |\hat{\mathbf{M}}\rho_{z,jk}(l)|, \quad 1 \leq j < k \leq m. \quad (3.5)$$

A total of  $d = m(m-1)/2$  values for the maximum moving cross-correlation statistics are needed to be calculated and then reordered in decreasing order as  $\hat{\mathbf{L}}_1 \geq \hat{\mathbf{L}}_2 \geq \dots \geq \hat{\mathbf{L}}_d$ .

We use the following ratio criterion to reject the hypothesis in (3.4) based on the statistics calculated in (3.5). Consider only the correlation corresponding to the largest  $\hat{r}$  maximum statistics in (3.5) to be significantly correlated, where  $\hat{r}$  is defined as

$$\hat{r} = \arg \max_{1 \leq i \leq hd} \hat{\mathbf{L}}_i / \hat{\mathbf{L}}_{i+1} \quad (3.6)$$

where  $h$  is a pre-specified number and  $0 < h < 1$ . A value of  $h = 0.75$  will be used as recommended by Chang et al. (2018) to avoid extreme cases, where the dominator in (3.6) is extremely close to zero (i.e.  $\hat{\mathbf{L}}_{i+1} \simeq 0$ ). Therefore, we report that the two components corresponding to  $\hat{\mathbf{L}}_j$  are significantly correlated, where  $1 \leq j \leq r$ . Notice that, using  $h = 0.75$  is one drawback of this method. A small amount of information might be lost because of using  $h$ , especially for extreme cases where strong correlations exist among all tested components. For these extreme cases, it is recommended to validate the results of the maximum moving cross-covariance method with those of the moving cross-correlation plots.

Therefore, the maximum moving cross-correlation method can be applied to find the  $n$  correlated components of  $\hat{\mathbf{f}}_t$  and hence correctly rearrange its columns to find the hidden segmentation of  $\hat{\mathbf{x}}_t$  as summarized below:

1. Calculate  $\hat{\mathbf{L}}_d(j, k)$  for the  $d = \frac{m(m-1)}{2}$  pairs of the components of  $\hat{\mathbf{f}}_t$ .
2. Rearrange the values of the calculated  $d$  statistics in step 1 from largest to smallest (i.e.  $\hat{\mathbf{L}}_1 \geq \hat{\mathbf{L}}_2 \geq \dots \geq \hat{\mathbf{L}}_d$ ).
3. Find the value of  $\hat{r}$  using the criterion in (3.6).
4. Consider the pair corresponding to the statistics  $\hat{\mathbf{L}}_j$  to be significantly correlated (i.e. reject the corresponding null hypothesis), where  $1 \leq j \leq \hat{r}$ .
5. Rearrange the columns of  $\hat{\mathbf{f}}_t$  by grouping correlated components according to step 4. The final result of this step is the vector with the hidden segmentation,  $\hat{\mathbf{x}}_t$ .

These steps can be considered as sub-steps of step 5 of the GTS-PCA procedure.

**Example 1.** This example is a short simulation study to test the ability of the maximum moving cross-correlation method to regroup the variables of a non-stationary time series based on their dynamic relationships. The simulated data with eight variables and 1200 observations in Example 1 of Alshammri and Pan (2019) are used. This series consists of 3 subseries of 4, 3 and 1 variable such that each subseries was generated using different non-stationary model as follows.

Let  $\nabla$  be the differencing notation, where  $\nabla \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ . Let  $a_t$ ,  $b_t$ , and  $c_t$  be three independent standard normal white noises, which are the innovation terms of the following three models, respectively, then:

$$\begin{cases} y_{j,t} = u_{t+j-1}, & j = 1, 2, 3 \text{ and } 4 \\ y_{j,t} = v_{t+j-4}, & j = 5, 6 \text{ and } 7 \\ y_{j,t} = w_t, & j = 8 \end{cases} \quad (3.7)$$

where  $(u_t, v_t, w_t)'$  satisfies

$$\begin{cases} \nabla u_t = 0.6 \nabla u_{t-1} + a_t + 0.5a_{t-1}, \\ \nabla v_t = b_t + 0.8b_{t-1} - 2.2b_{t-2}, \\ \nabla w_t = -0.55 \nabla w_{t-1} - 0.5 \nabla w_{t-2} + 0.4 \nabla w_{t-3} + c_t. \end{cases} \quad (3.8)$$

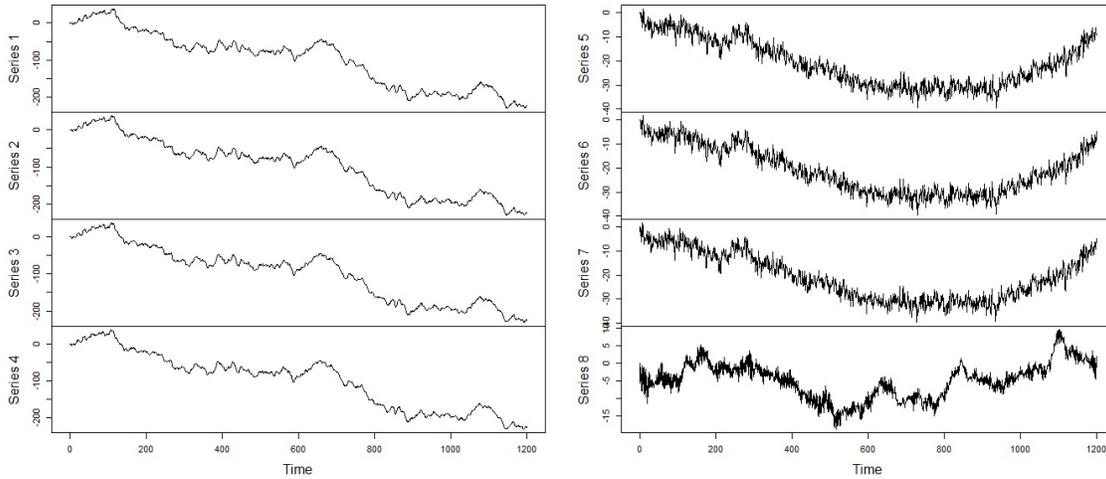


Figure 1: A time series plot of the simulated series with 8 variables.

Results Method	Correlated Variables
Maximum Moving Cross-Correlation (window=101)	{1, 2, 3, 4}, {5, 6, 7} and {8}
Moving Cross-Correlation (window=101)	{1, 2, 3, 4}, {5, 6, 7} and {8}

Table 1: Correlated variables of the simulated data with 8 variables based on different methods.

A time series plots of the simulated data is shown below in Figure 1. A window with size 101 is used in the calculation of the moving cross-correlation function. The outcomes of the maximum moving cross-correlation method are reported and validated with those of moving cross-correlation method of Alshammri and Pan (2019); See Table 1. Both methods were able to capture the dynamic relationship among different simulated variables where 3 uncorrelated subseries of 4, 3 and 1 variable are suggested.

## 4 Theoretical Properties

For the GTS-PCA's model in equation (2.1), we will show that the transformation matrix  $\hat{\mathbf{A}}$  obtained above is a consistent estimator of  $\mathbf{A}$ . To this end, we are going to use the following distance measurement. Let  $c, c_1, c_2, \dots$  denote constants whose values might differ

from place to place. For  $c_1 < c_2$ , let  $\mathbf{B}_1$  and  $\mathbf{B}_2$  be any  $c_2 \times (c_2 - c_1)$  matrices satisfying the condition  $\mathbf{B}'_i \mathbf{B}_i = \mathbf{I}_{(c_2 - c_1) \times (c_2 - c_1)}$ , where  $i \in \{1, 2\}$ , then the distance between the  $\mathbf{B}_1$  and  $\mathbf{B}_2$  is

$$D(\mathbf{B}_1, \mathbf{B}_2) = \sqrt{c_2 - c_1 - \text{tr}(\mathbf{B}_1 \mathbf{B}'_1 \mathbf{B}_2 \mathbf{B}'_2)}. \quad (4.1)$$

In particular, let  $W = 2w + 1$  be the window size, then we shall approach the consistency by showing that  $D(\mathcal{M}(\hat{\mathbf{A}}), \mathcal{M}(\mathbf{A})) \rightarrow 0$  as  $W \rightarrow \infty$ , where  $\mathcal{M}(\mathbf{A})$  is the linear space spanned by  $\mathbf{A}$ 's columns and  $D(\mathcal{M}(\hat{\mathbf{A}}), \mathcal{M}(\mathbf{A}))$  is the distance between the spaces  $\mathcal{M}(\hat{\mathbf{A}})$  and  $\mathcal{M}(\mathbf{A})$ . The measurement in (4.1) was also used by Chang et al. (2018) and Alshammri and Pan (2019).

The convergence of  $\hat{\mathbf{A}}$  can be implied mainly by the convergence of  $\hat{\mathbf{M}}\Gamma_z(l)$ , which is used to calculate  $\hat{\mathbf{M}}\hat{\mathbf{W}}_z$  to finally formulate  $\hat{\mathbf{U}}_z$ . Recall that  $\hat{\mathbf{U}}_z$  is equivalent to  $\hat{\mathbf{A}}$  after appropriate arrangement of its columns. Hence, we shall assume that  $\hat{\mathbf{U}}_z$  has a correct columns rearrangement and therefore  $\hat{\mathbf{U}}_z = \hat{\mathbf{A}}$ . This assumption is justified as the order of the columns will not affect the spanned linear space. In this section, we are going to show the consistency of our estimator when  $m$  is fixed and when the growth rate of  $m$  depends on  $W$ . Also, we are going to consider the following measurement of dependence:

$$\theta_l = \sup_c \sup_{A \in \mathcal{F}_{-\infty}^c, B \in \mathcal{F}_{c+l}^\infty} |P(A \cap B) - P(A)P(B)| \quad (4.2)$$

where  $\mathcal{F}_{c_3}^{c_4}$  is the  $\sigma$ -field generated by  $\mathbf{z}_t$  for  $c_3 \leq t \leq c_4$ . The measurement of dependence in (4.2) is also called the mixing coefficients.

We are going to assume that the transformed  $n$  blocks of  $\mathbf{M}\mathbf{W}_x$  do not share the same eigenvalues, otherwise these blocks should be merged as one block. In other words, denote by  $\sigma(\mathbf{M}\mathbf{W}_x)$  the set of all eigenvalues of  $\mathbf{M}\mathbf{W}_x$ , then the minimum distance between any two eigenvalues from the different blocks  $\lambda_j, \lambda_k \in \sigma(\mathbf{M}\mathbf{W}_x)$

$$d_W = \min_{1 \leq j < k \leq n} |\lambda_j - \lambda_k| \quad (4.3)$$

is positive.

#### 4.1 Asymptotic Results ( $m$ is fixed)

For ease of notation, assume  $\max(l, w) = w$ . We need some regularity assumptions.

**Assumption 1.** Assume

$$\sup_i \max_{1 \leq j \leq m} E|z_{j,i} - E(z_{j,i})|^{2q}$$

to be upper bounded by a positive constant  $M$  for some constant  $q > 2$ , where  $z_{j,i}$  is the  $j^{\text{th}}$  component of  $\mathbf{z}_i$ .

**Assumption 2.** Assume that the mixing coefficients  $\theta_l$  defined in (4.2) satisfies  $\sum_{l=1}^{\infty} \theta_l^{1-2/q} < \infty$  for  $q$  defined in Assumption 1.

**Assumption 3.** Assume that different blocks of  $\mathbf{M}\mathbf{W}_x$  do not share the same eigenvalues. That is  $d_W > 0$ , where  $d_W$  is defined in (4.3).

**Theorem 1.** Under assumptions 1-3 and for a fixed dimension  $m$ , there exists  $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \dots, \hat{\mathbf{A}}_n)$  such that, as  $W \rightarrow \infty$ ,

$$\max_{1 \leq j \leq n} D(\mathcal{M}(\hat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j)) \xrightarrow{P} 0.$$

The proof of Theorem 1 is given in the Appendix.

## 4.2 Asymptotic Results ( $m = o(W^c)$ )

As shown earlier, when the dimension  $m$  is fixed, the sample moving cross-covariance matrix  $\hat{\mathbf{M}}\Gamma_z(l)$  defined in (2.5) is convergent with rate  $W^{-1/2}$ . However, for the case where  $m$  is large and its growth rate depends on  $W$ ,  $\hat{\mathbf{M}}\Gamma_z(l)$  is no longer convergent. This is because each of the sample cross-covariance matrices used on each window is not consistent; See Chen et al. (2013). One solution, in this case, is to use a threshold estimator; See Bickel and Levina (2008). For the GTS-PCA, the threshold estimator can be applied to the sample cross-covariance matrix  $\hat{\Gamma}_{z,i}(l)$  in order to obtain a consistent estimator for  $\Gamma_{z,i}(l)$ . The obtained new matrices then can be used in the calculation of the sample moving cross-covariance matrix.

Specifically, write the sample lag  $l$  cross-covariance matrix over window  $\mathbf{z}_i$  defined in (2.3) as  $\hat{\Gamma}_{z,i}(l) = (\hat{s}_{j,k}^{(i)}(l))_{1 \leq j,k \leq m}$ . Write the sample lag  $l$  moving cross-covariance matrix defined in (2.5) as  $\hat{\mathbf{M}}\Gamma_{z,i}(l) = (\hat{\sigma}_{j,k}(l))_{1 \leq j,k \leq m}$ . Then, the threshold estimator of the sample cross-covariance matrix of window  $\mathbf{z}_i$  is defined as

$$T_u(\hat{\Gamma}_{z,i}(l)) = (\hat{s}_{j,k}^{(i)}(l) I\{|\hat{s}_{j,k}^{(i)}(l)| \geq u\})_{1 \leq j,k \leq m}, \quad (4.4)$$

where  $I\{\cdot\}$  is the indicator function. When the threshold level  $u$  is chosen to be  $u = cm^{2/\gamma} W^{-1/2}$ , and under some regularity conditions for a specified constant  $\gamma$  and a positive

constant  $c$ , then  $\max_{j,k} |\hat{s}_{j,k}^{(i)}(l) - s_{j,k}^{(i)}(l)| = O_p(m^{2/\gamma} W^{-1/2})$ ; See Chang et al. (2018). Notice that  $\hat{s}_{j,k}^{(i)}(l)$  now are the elements of the threshold estimator defined in (4.4).

The sample lag  $l$  moving cross-covariance matrix of  $\mathbf{z}_t$  defined in (2.5) can be redefined as

$$\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) = \frac{1}{T - 2 \max(l, w)} \sum_{i=\max(l,w)+1}^{T-\max(l,w)} T_u(\hat{\Gamma}_{z,i}(l)). \quad (4.5)$$

$\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l)$  in this case will be a consistent estimator of  $\mathbf{M}\Gamma_z(l)$ ; See Lemma 5 in the appendix. Also, redefine the sample quadratic order of moving cross-covariance matrix of  $\mathbf{z}_t$  in (2.10) as

$$\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} = \mathbf{I}_m + \sum_{l=1}^{l_1} \hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) \cdot [\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l)]'. \quad (4.6)$$

Therefore, the calculation procedure of GTS-PCA in the case where  $m = o(W^c)$  will remain the same with replacing  $\hat{\mathbf{M}}\mathbf{W}_z$  by  $\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}}$ . Under stronger versions of assumptions 1 and 2, the consistency of  $\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}}$  will be shown in this section.

Define the quadratic order cross-covariance matrix over window  $\mathbf{z}_i$  to be

$$\mathbf{W}\Gamma_{z,i} = \mathbf{I}_m + \sum_{l=1}^{l_1} \Gamma_{z,i}(l) \cdot [\Gamma_{z,i}(l)]' \quad (4.7)$$

which can be estimated by

$$\hat{\mathbf{W}}\Gamma_{z,i}^{\text{thre}} = \mathbf{I}_m + \sum_{l=1}^{l_1} T_u(\hat{\Gamma}_{z,i}(l)) \cdot [T_u(\hat{\Gamma}_{z,i}(l))]' \quad (4.8)$$

Let  $\mathbf{A}^{(i)}$  be the matrix that consists of the eigenvectors of  $\mathbf{W}\Gamma_{z,i}$  defined in (4.7), where  $\mathbf{A}^{(i)} = (a_{j,k}^{(i)})_{1 \leq j,k \leq m}$ . In the stationary TS-PCA,  $\mathbf{A}^{(i)}$  is used to transform an  $m$ -dimensional series into uncorrelated  $r$  subseries, where the window  $\mathbf{z}_i$  consists of all data in the series (i.e. only one window is used as  $W = T$ ). Denote by

$$B = \max_{1 \leq k \leq r} (m_k) \quad (4.9)$$

the maximum dimension among the  $r$  different subseries transformed by  $\mathbf{A}^{(i)}$  after columns' rearrangement.

In what follows, we will show the consistency of  $\hat{\mathbf{A}}$  based on the threshold estimators, as defined in (4.4), (4.5) and (4.6). In particular, we shall state the following assumptions.

**Assumption 4.** Assume for each window  $\mathbf{z}_i$  that

$$\max_{1 \leq k \leq m} \sum_{j=1}^m |a_{j,k}^{(i)}|^\delta \leq c_5 \quad (4.10)$$

and

$$\max_{1 \leq j \leq m} \sum_{k=1}^m |a_{j,k}^{(i)}|^\delta \leq c_6$$

for an arbitrary constant  $\delta \in [0, 1)$ , where  $c_5$  and  $c_6$  are positive and allowed to diverge away with  $m$ .

Assumption 4 controls the sparsity of the matrix  $\mathbf{A}^{(i)}$  through  $\delta, c_5$  and  $c_6$  when the dimension  $m$  is large.  $\hat{\mathbf{A}}^{(i)}$  converges to  $\mathbf{A}^{(i)}$  faster when its sparsity level increases; See Chang et al. (2018).

**Assumption 5.** Assume that as  $g \rightarrow \infty$ ,

$$\sup_i \max_{1 \leq j \leq m} P(|z_{j,i} - E(z_{j,i})| > g) = O(g^{-2(\gamma+\varepsilon)})$$

for constants  $\gamma > 2$  and  $\varepsilon > 0$ .

Assumption 5 is a stronger version of Assumption 1, where the tail probabilities of  $\mathbf{z}_i$  is required to decay faster than those in Assumption 1.

**Assumption 6.** Assume that as  $l \rightarrow \infty$ ,

$$\theta_l = O(l^{-\gamma(\gamma+\varepsilon)/2\varepsilon})$$

where  $\theta_l$  is the mixing coefficients defined in (4.2), and the constants  $\gamma$  and  $\varepsilon$  are provided in Assumption 5.

Assumption 6 requires the mixing coefficients to decay faster than those in Assumption 2. Before stating the last assumption, consider the following notations. Let

$$f = Bc_5c_6 \tag{4.11}$$

where  $c_5, c_6$  are given in Assumption 4 and  $B$  is given in (4.9). Let

$$d_k = \min_{1 \leq j \leq n, \lambda_j \neq \lambda_k} |\lambda_j - \lambda_k| \tag{4.12}$$

for  $1 \leq k \leq n$ , and  $\lambda_j$  and  $\lambda_k$  are the eigenvalues of the  $j^{\text{th}}$  and  $k^{\text{th}}$  blocks of  $\mathbf{M}\mathbf{W}_x$ , respectively. Let

$$h = \max_{1 \leq l \leq l_1} \|\mathbf{M}\mathbf{\Gamma}_x(l)\|_2 \tag{4.13}$$

**Assumption 7.** Assume that  $d_k > 0$  for  $1 \leq k \leq n$ , where  $d_k$  is defined in (4.12).

**Theorem 2.** Under assumptions 4-7, there exists  $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \dots, \hat{\mathbf{A}}_n)$  such that

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) \xrightarrow{P} 0$$

hold for  $m = o(W^{\gamma/4})$ , as  $W \rightarrow \infty$ .

The proof of Theorem 2 is provided in the Appendix.

### 4.3 Asymptotic Results ( $\log m = o(W^c)$ )

We now study the asymptotic properties of the transformation matrix  $\hat{\mathbf{A}}$  in the case where the growth rate of  $m$  has an exponential relationship with  $W$ . The sample cross-covariance matrices that are calculated over the windows will not be convergent in this case. We shall apply a threshold estimator to  $\Gamma_z(l)$ . In particular, the threshold estimator  $T_u(\hat{\Gamma}_{z,i}(l))$  defined in (4.4) will be used, but with a different threshold level than the previously used. Chang et al. (2018) showed that when  $\log m = o(W^c)$ , under some regularity assumptions, the threshold estimator of the cross-covariance matrix with a threshold level  $u = c(W^{-1} \log m)^{1/2}$  is consistent. This result can be applied to  $T_u(\hat{\Gamma}_{z,i}(l))$  over each window to calculate  $\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l)$  as follows.

**Assumption 8.** Assume for any  $g > 0$  and  $\|\mathbf{y}\|_2 = 1$  that

$$\sup_i P(|\mathbf{y}'(\mathbf{z}_i - E(\mathbf{z}_i))| > g) \leq c_7 \exp(-c_8 g^\eta)$$

where the constants  $c_7, c_8$  are positive and  $\eta \in (0, 2]$ .

This assumption implies that the tail probabilities of linear combinations of  $\mathbf{z}_i$  decays exponentially.

**Assumption 9.** Assume that

$$\theta_l \leq \exp(-c_9 l^\xi)$$

for all  $l \geq 0$ , where  $\xi \in (0, 1]$  and  $c_9$  is a positive constant.

This assumption requires the mixing coefficient to decay exponentially.

**Theorem 3.** Suppose assumptions 4, 7, 8 and 9 hold. Let  $\log m = o(W^{b/(2-b)})$ , where  $b = 1/(2\eta^{-1} + \xi^{-1})$ . Then there exists  $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \dots, \hat{\mathbf{A}}_n)$  such that

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ .

The proof of Theorem 3 is available in the Appendix.

## 5 Simulations and Real Data Examples

This section provides the numerical properties of GTS-PCA. Our method will be tested on both simulated and real data sets with various dimensions and sample sizes. All analyses are carried out using R software. The necessary functions to produce and assess the results of GTS-PCA can be found under the following R packages: `GTSPCA` of Alshammri (2020a) and `MpermutMax` of Alshammri (2020b).

### 5.1 Simulation

The simulated data are generated using the following steps. First, we generate the subseries of  $\mathbf{x}_t$  separately using the `arma.sim` command. Then, the elements of the transformation matrix  $\mathbf{A}$  are generated randomly from  $unif(-5, 5)$  using the `runif` command. Post multiplying  $\mathbf{x}_t$  by  $\mathbf{A}$  will generate  $\mathbf{z}_t$  as in (2.1). GTS-PCA with different window sizes  $W = \{101, 201, 501\}$  and TS-PCA (i.e. GTS-PCA with  $W = T$ ) will be applied to simulated data. Then, the percentage of correct segmentation will measure the accuracy of the results, where *correct segmentation* refer to the case where the final transformed series is correctly rearranged into uncorrelated subseries as simulated in  $\mathbf{x}_t$  at the first step of the simulation process. For demonstration purposes, a choice of five-time lags (i.e.  $l_1 = 5$ ) will be used in the analyses unless stated otherwise.

**Example 2.** The first simulation study includes a non-stationary time series  $\mathbf{z}_t$  with a dimension of  $m = 5$ . Three different models are used to generate the subgroups (i.e. subseries) of this series, which consists of 2, 2 and 1 independent variables.

The subgroups of  $\mathbf{x}_t$  are simulated as follows. Let  $a_t$ ,  $b_t$  and  $c_t$  be independent standard normal white noises, which are the innovation terms of the following three models, respectively, then:

$$\begin{cases} x_{j,t} = u_{t+j-1}, & j = 1 \text{ and } 2 \\ x_{j,t} = v_{t+j-3}, & j = 3 \text{ and } 4 \\ x_{j,t} = w_t, & j = 5 \end{cases} \quad (5.1)$$

where  $(u_t, v_t, w_t)'$  satisfies

$$\begin{cases} \nabla u_t = 0.75 \nabla u_{t-1} + a_t + 1.5a_{t-1} + 0.5a_{t-2} + 0.2a_{t-3} \\ \nabla v_t = -0.5 \nabla v_{t-1} - 0.4 \nabla v_{t-2} + b_t - 1.2b_{t-1} - 2.75b_{t-2} + 0.5b_{t-3} \\ w_t = -0.9w_{t-1} - 0.45w_{t-2} + c_t - 1.1c_{t-1} - 0.8c_{t-2} - 0.6c_{t-3} - 1.3c_{t-4} \end{cases} \quad (5.2)$$

Then, we post multiply  $\mathbf{x}_t$  by  $\mathbf{A}$  whose elements are generated randomly from  $unif(-5, 5)$  to obtain  $\mathbf{z}_t$ , as mentioned earlier. GTS-PCA with different window sizes is applied to the simulated series.

Method	GTS-PCA with W=101						
	T=300	T=400	T=500	T=600	T=1000	T=1500	T=2000
Percentage of Correct Segmentation	0.33	0.51	0.54	0.64	0.86	0.94	0.96
Method	GTS-PCA with W=201						
	T=300	T=400	T=500	T=600	T=1000	T=1500	T=2000
Percentage of Correct Segmentation	0.15	0.18	0.29	0.38	0.55	0.72	0.75
Method	GTS-PCA with W=501						
	T=300	T=400	T=500	T=600	T=1000	T=1500	T=2000
Percentage of Correct Segmentation	NA	NA	NA	0.07	0.15	0.24	0.28
Method	TS-PCA						
	T=300	T=400	T=500	T=600	T=1000	T=1500	T=2000
Percentage of Correct Segmentation	0.44	0.47	0.52	0.49	0.47	0.50	0.43

Table 2: Percentages of correct segmentation of 500 replication of the simulated series with 5 variables.

The above simulation was replicated 500 times, and the percentage of correct segmentations are reported in Table 2. It can be seen that the performance of GTS-PCA was enhanced by increasing  $T$ . GTS-PCA with small window sizes (e.g.  $W = 101$  and  $W = 201$ ) produced more accurate results compared with those with larger windows (e.g.  $W = 501$ ). This is mainly because of the non-stationarity of the data. TS-PCA, on the other hand, produced the correct segmentations about 47% of the time for different sizes of  $T$ . Hence, GTS-PCA with a small window size is more suitable to capture the hidden

segmentation in this example.

GTS-PCA with  $W = 101$  is of particular interest, as it outperformed those with larger window sizes. It produced reasonable results for moderate sample sizes (e.g.  $T \geq 600$ ) in this example. For instance, it produced the correct segmentations 86% of the time when  $T = 1000$ , and 96% of the time when  $T = 2000$ .

Consider one randomly selected replica of the above simulations of  $\mathbf{z}_t$  with  $T = 1000$ . A time series plot of  $\mathbf{z}_t$  is shown in Figure 2. All five variables exhibited similar non-stationary behaviours. Sample moving cross-correlation plots with  $W = 101$  between the variables of  $\mathbf{z}_t$  before transformation are shown in Figure 3, where all variables are correlated. Figure 4 provides moving cross-correlation plots with  $W = 101$  between the components of  $\mathbf{z}_t$  after transformation by GTS-PCA with  $W = 101$ . Notice that the transformed series consisted of three uncorrelated subseries, namely  $\{1, 3\}$ ,  $\{2, 5\}$  and  $\{4\}$ , which revealed the hidden correct segmentation of  $\mathbf{x}_t$  with 2, 2 and 1 variable.

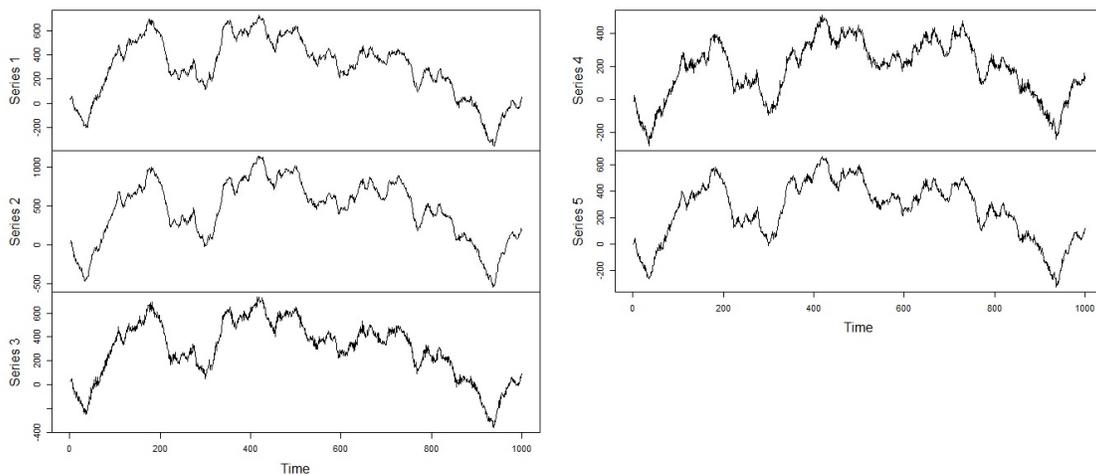


Figure 2: A time series plot of the simulated  $\mathbf{z}_t$  with 5 variables.

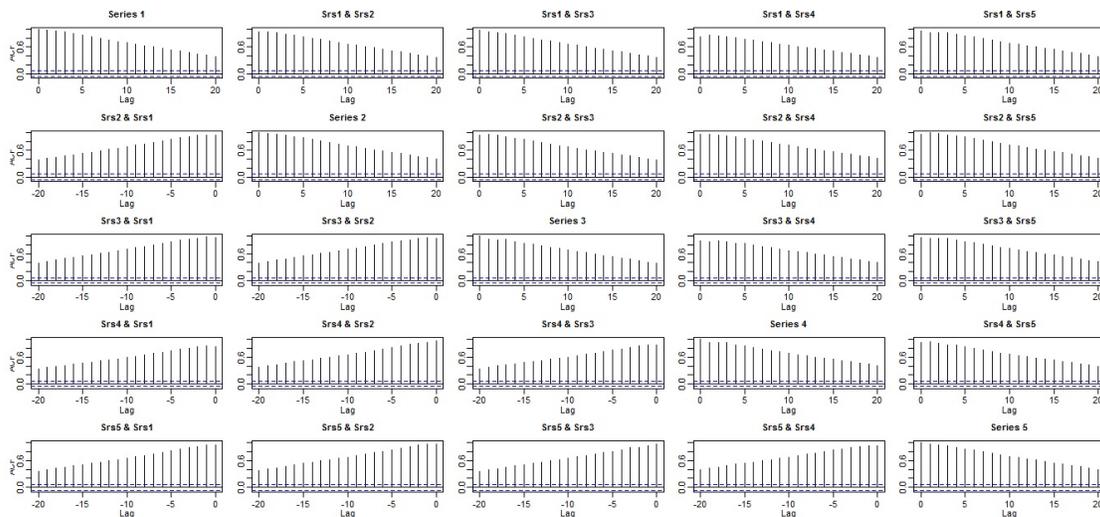


Figure 3: Sample moving cross-correlation plots with  $W = 101$  between the components of the simulated  $\mathbf{z}_t$  with 5 variables before transformation.

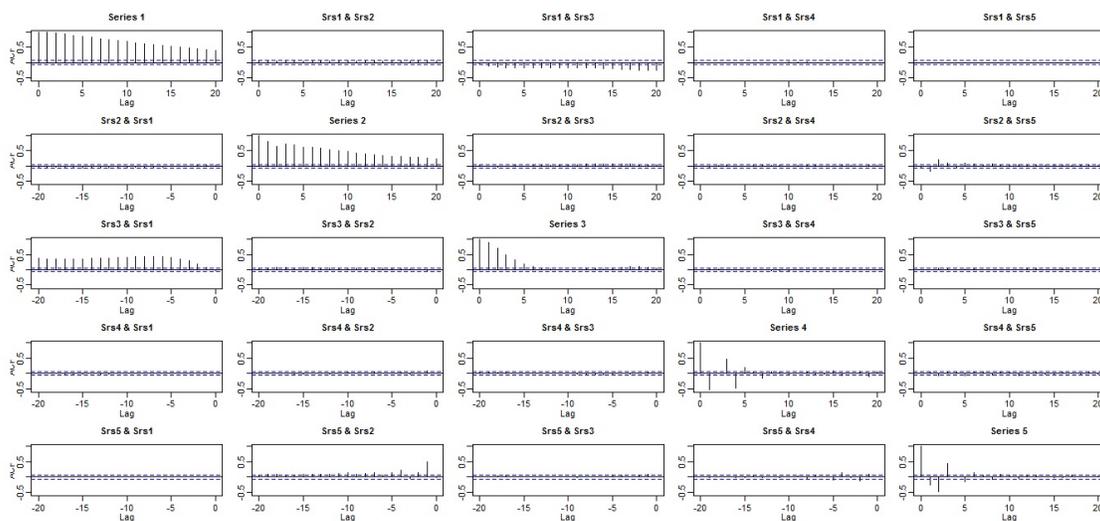


Figure 4: Sample moving cross-correlation plots with  $W = 101$  between the components of the simulated  $\mathbf{z}_t$  with 5 variables after GTS-PCA with  $W = 101$  is applied.

**Example 3.** In this example, a non-stationary time series with a dimension of  $m = 20$  is simulated such that it consists of five uncorrelated subseries with 6, 5, 4, 3 and 2 correlated variables, respectively.

The five subseries of  $\mathbf{x}_t$  are simulated as follows. Let  $a_t, b_t, c_t, d_t$  and  $e_t$  be independent standard normal white noises, which are the innovation terms of the following five models,

respectively, then:

$$\begin{cases} x_{j,t} = u_{t+j-1}, & j = 1, 2, 3, 4, 5 \text{ and } 6 \\ x_{j,t} = v_{t+j-7}, & j = 7, 8, 9, 10 \text{ and } 11 \\ x_{j,t} = w_{t+j-12}, & j = 12, 13, 14 \text{ and } 15 \\ x_{j,t} = f_{t+j-16}, & j = 16, 17 \text{ and } 18 \\ x_{j,t} = g_{t+j-19}, & j = 19 \text{ and } 20 \end{cases} \quad (5.3)$$

where  $(u_t, v_t, w_t, f_t, g_t)'$  satisfies

$$\begin{cases} \nabla u_t = 0.8 \nabla u_{t-1} + a_t + 3a_{t-1} \\ v_t = -0.9v_{t-1} - 0.4v_{t-2} + b_t - b_{t-1} - 0.8b_{t-2} - 0.6b_{t-3} - 1.4b_{t-4} \\ \nabla w_t = -0.55 \nabla w_{t-1} + c_t - 2.1c_{t-1} \\ \nabla f_t = -0.5 \nabla f_{t-1} - 0.4 \nabla f_{t-2} + d_t - 1.2d_{t-1} - 2.75d_{t-2} + 0.5d_{t-3} \\ \nabla g_t = 0.4 \nabla g_{t-1} - 0.3 \nabla g_{t-2} + e_t + 2.9e_{t-1} - 0.8e_{t-2} - 1.8e_{t-3} - 1.9e_{t-4} \end{cases} \quad (5.4)$$

Method	GTS-PCA with W=101						
	T=1500	T=2000	T=2500	T=3000	T=3500	T=4000	T=4500
Percentage of Correct Segmentation	0.13	0.35	0.59	0.78	0.84	0.92	0.95
Method	GTS-PCA with W=201						
	T=1500	T=2000	T=2500	T=3000	T=3500	T=4000	T=4500
Percentage of Correct Segmentation	0.11	0.21	0.28	0.30	0.39	0.42	0.43
Method	GTS-PCA with W=501						
	T=1500	T=2000	T=2500	T=3000	T=3500	T=4000	T=4500
Percentage of Correct Segmentation	0.01	0.10	0.12	0.13	0.15	0.16	0.17
Method	TS-PCA						
	T=1500	T=2000	T=2500	T=3000	T=3500	T=4000	T=4500
Percentage of Correct Segmentation	0.00	0.02	0.01	0.01	0.01	0.02	0.03

Table 3: Percentages of correct segmentation of 500 replication of the simulated series with 20 variables.

The series  $\mathbf{z}_t$  then is generated as in the previous example. Based on 500 replications of the above simulation, Table 3 summarizes the obtained percentages of correct segmentations. The accuracy of GTS-PCA improved as the sample size is increased, especially for small window sizes, where the GTS-PCA captured more information out of the non-stationary series. For example, GTS-PCA with  $W = 101$  produced the correct segmentations 78% of the time when  $T = 3000$ , and 95% of the time when  $T = 4500$ . Furthermore, the performance of GTS-PCA decreased as we increased  $W$ . For instance, when  $T = 3500$ , GTS-PCA with windows sizes 101, 201 and 501 produced the correct segmentations 84%, 30% and 13% of the time, respectively. On the other hand, TS-PCA (i.e. GTS-PCA with  $W = T$ ) had a poor performance on the simulated non-stationary series with percentages of correct segmentations less than 3 % in all cases. Therefore, GTS-PCA with  $W = 101$  is suggested to reduce the dimension of the simulated series in this example.

Based on one randomly selected replica with  $T = 3000$ , a time series plot of  $\mathbf{z}_t$  is shown in figures 5 and 6. First and last 50 sample moving cross-correlation plots with  $W = 101$  between the components of  $\mathbf{z}_t$  before transformation are shown in figures 7, 8, 9 and 10. The plots suggested that strong relationships existed among the components of  $\mathbf{z}_t$ . After applying GTS-PCA with  $W = 101$  on  $\mathbf{z}_t$ , the maximum moving cross-correlation method with  $W = 101$  suggested to rearrange the transformed series into five uncorrelated subgroups that consist of the following variables, respectively:  $\{1, 2, 7, 11, 15, 20\}$ ,  $\{3, 6, 9, 13, 18\}$ ,  $\{4, 8, 14, 19\}$ ,  $\{5, 12, 17\}$  and  $\{10, 16\}$ , which match the hidden correct segmentations of  $\mathbf{x}_t$ .

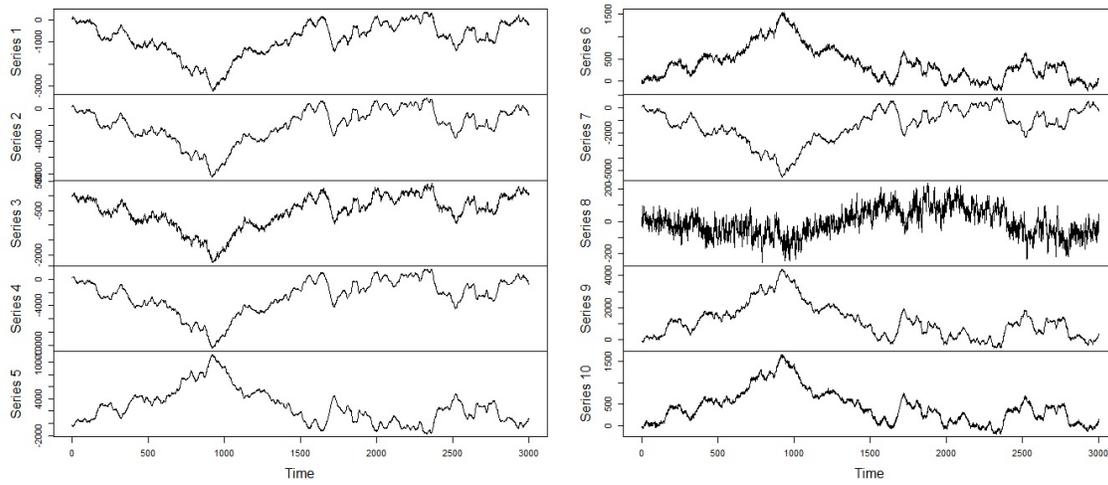


Figure 5: A time series plot of variables 1 - 10 of the simulated data with 20 variables.

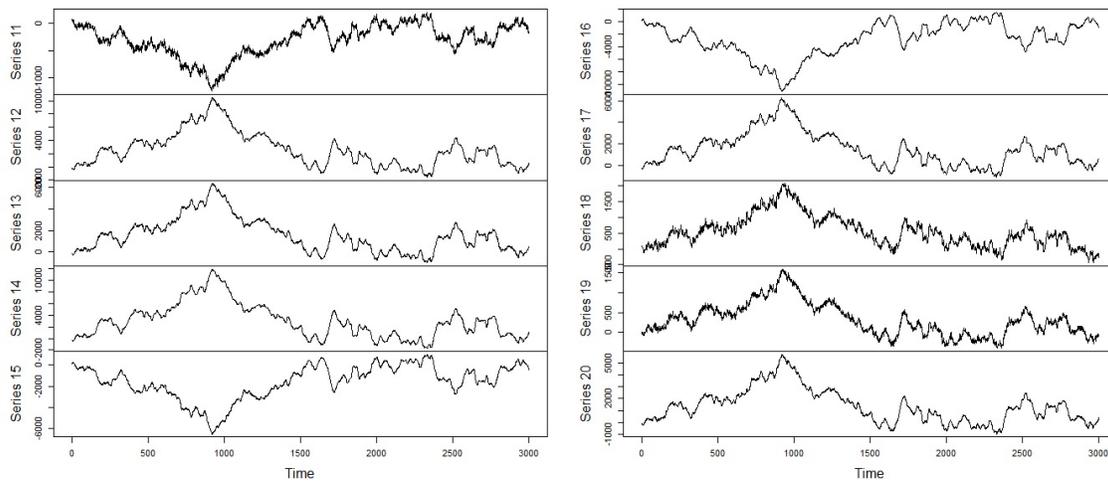


Figure 6: A time series plot of the variables 11 - 20 of the simulated data with 20 variables.

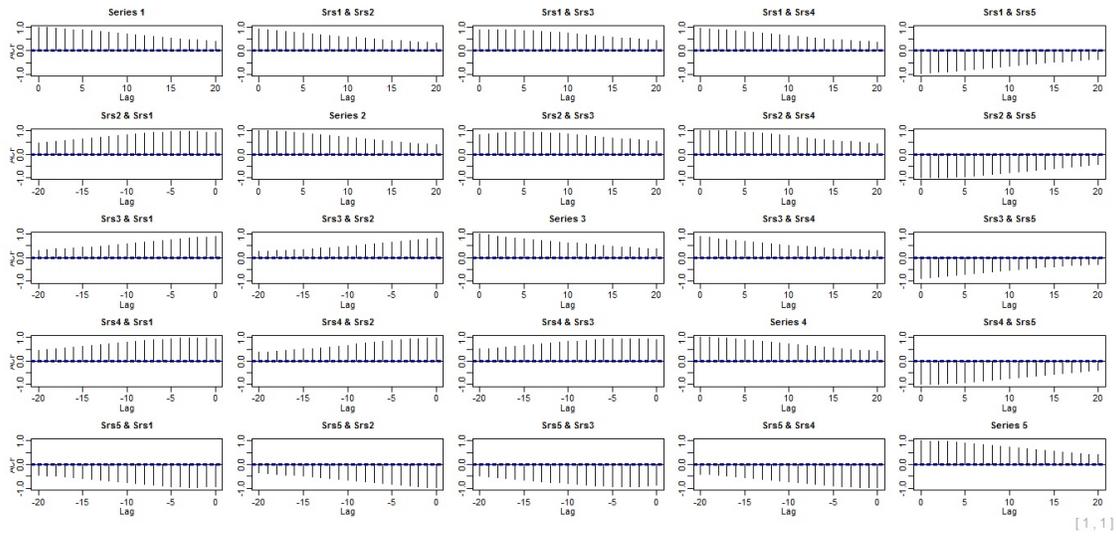


Figure 7: First 50 sample moving cross-correlation plots with window size 101 for the simulated data before before transformation - part 1.

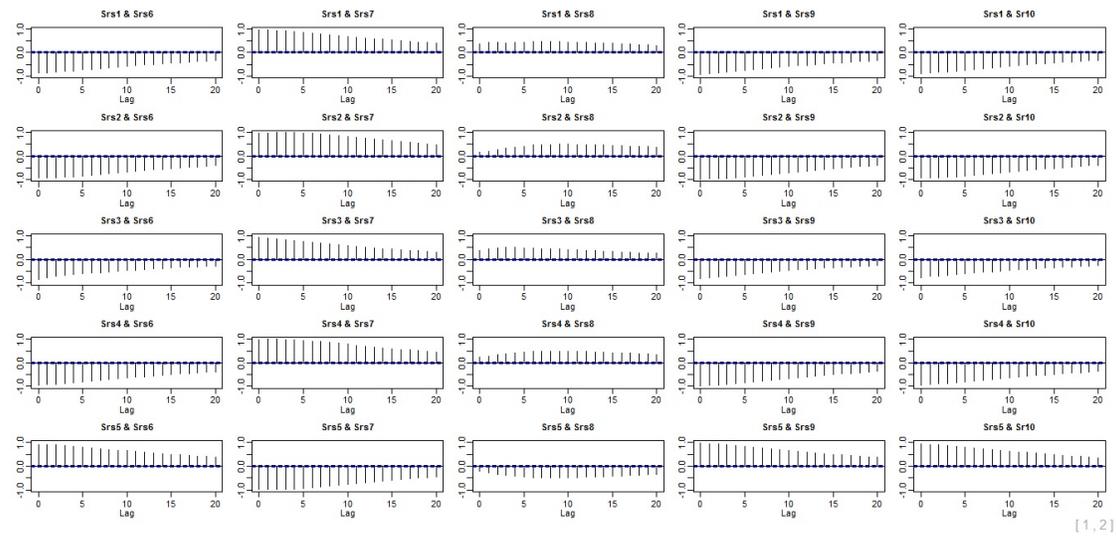


Figure 8: First 50 sample moving cross-correlation plots with window size 101 for the simulated data before before transformation - part 2.

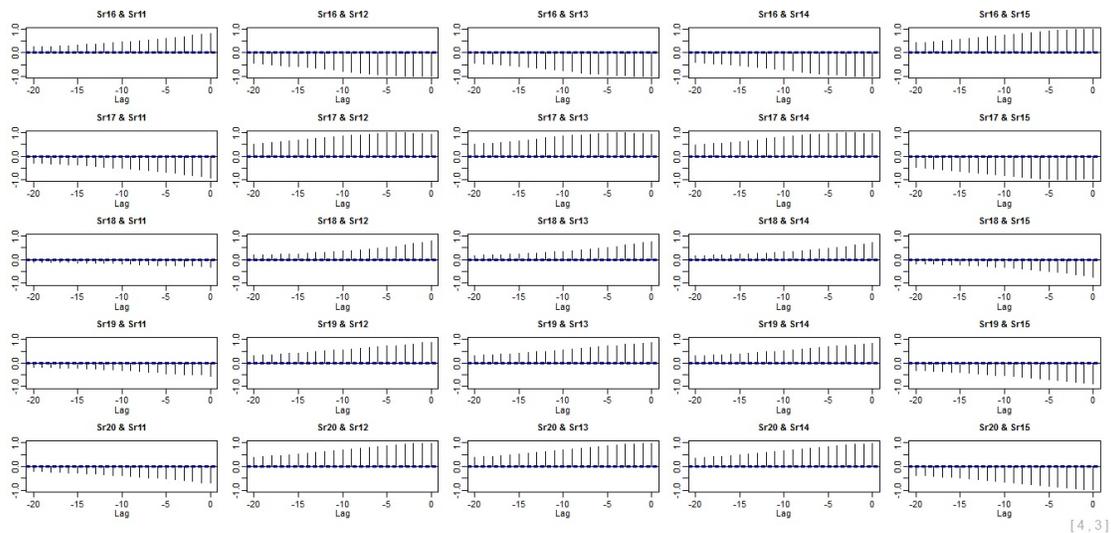


Figure 9: Last 50 sample moving cross-correlation plots with window size 101 for the simulated data before before transformation - part 1.

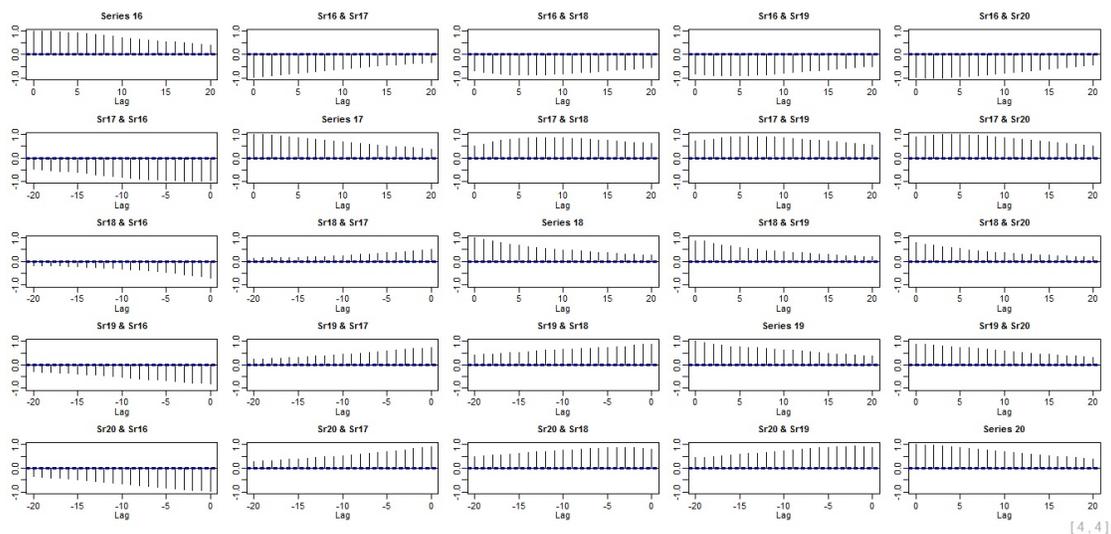


Figure 10: Last 50 sample moving cross-correlation plots with window size 101 for the simulated data before before transformation - part 2.

In conclusion, the dynamic relationships between the 20 non-stationary series were successfully captured by GTS-PCA with  $W = 101$  and the dimension was successfully reduced to five uncorrelated subseries of 6, 5, 4, 3 and 2 variables. Notice that, since we increased the dimension of the simulated series to  $m = 20$  in this example, larger sample sizes were needed to obtain accurate results comparing with those in the previous example.

**Example 4.** In this example, GTS-PCA is tested on a stationary data set, and then its results are validated with those of TS-PCA. Ideally, when GTS-PCA is applied to stationary data, then one window with a size  $W = T$  would be suggested. This is equivalent to use TS-PCA. However, one would ask about the behaviour of GTS-PCA when it uses a window size that is smaller than the ideal size. In this simulation study, we are going to answer this question. We will consider almost the worst-case scenario, where GTS-PCA with a small window size (e.g.  $W = 101$ ) is applied to a stationary data.

The data set used in this example is the one used by Chang et al. (2018) with  $m = 6$  and  $T = 1500$ . The simulation codes are available in the *PCA4TS* library in R software. This data set consists of three subgroups of dimensions 3, 2 and 1. A time series plot of the stationary series is available in Figure 11. Sample moving cross-correlation plots to show the relationships between the variables before transformation are provided in Figure 12. A moderate correlation existed across all six variables. A comparison between the results of GTS-PCA with  $W = 101$  and TS-PCA are summarized in Table 4. Sample moving cross-correlation plots of the data after transformation are shown in Figure 13. GTS-PCA and TS-PCA produced the same results for this data set, as they were both able to successfully find the correct segmentations.

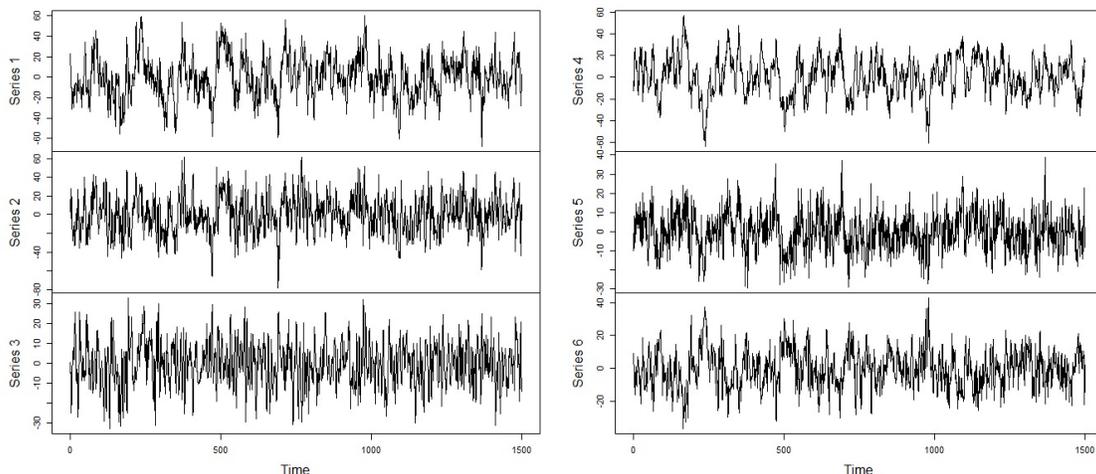


Figure 11: A time series plot of the simulated data with 6 stationary variables.

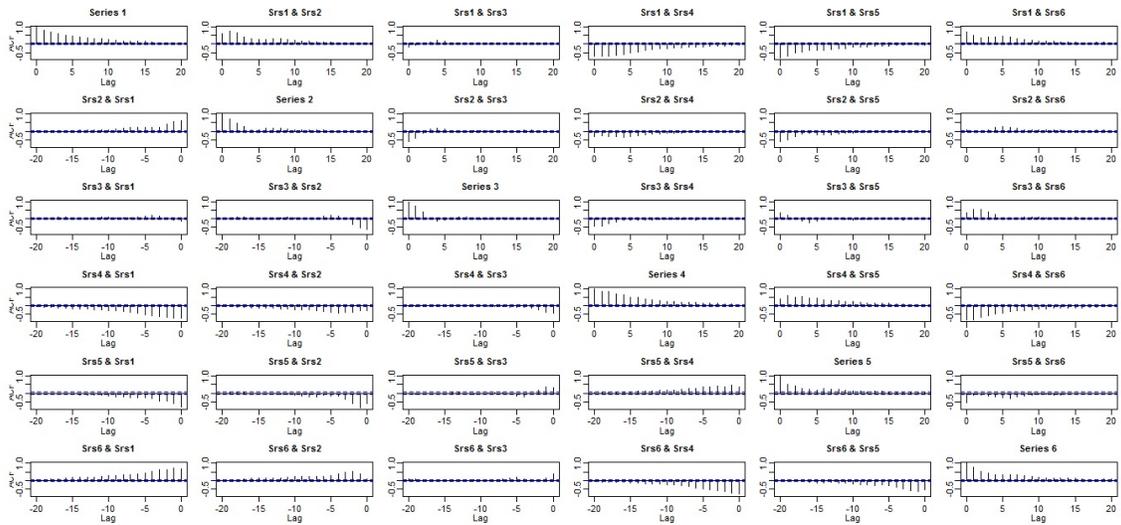


Figure 12: Sample moving cross-correlation plots with window size 101 for the simulated data before transformation.

Stationary Series $m=6, T=1500$	Number of subseries	Variables in subseries
GTS-PCA (window =101)	3 subseries of sizes 3, 2, 1	$\{1, 3, 6\}, \{2, 4\}, \{5\}$
TS-PCA	3 subseries of sizes 3, 2, 1	$\{1, 3, 6\}, \{2, 4\}, \{5\}$

Table 4: Number of subseries after applying GTS-PCA and TS-PCA to the simulated data with 6 stationary variables.

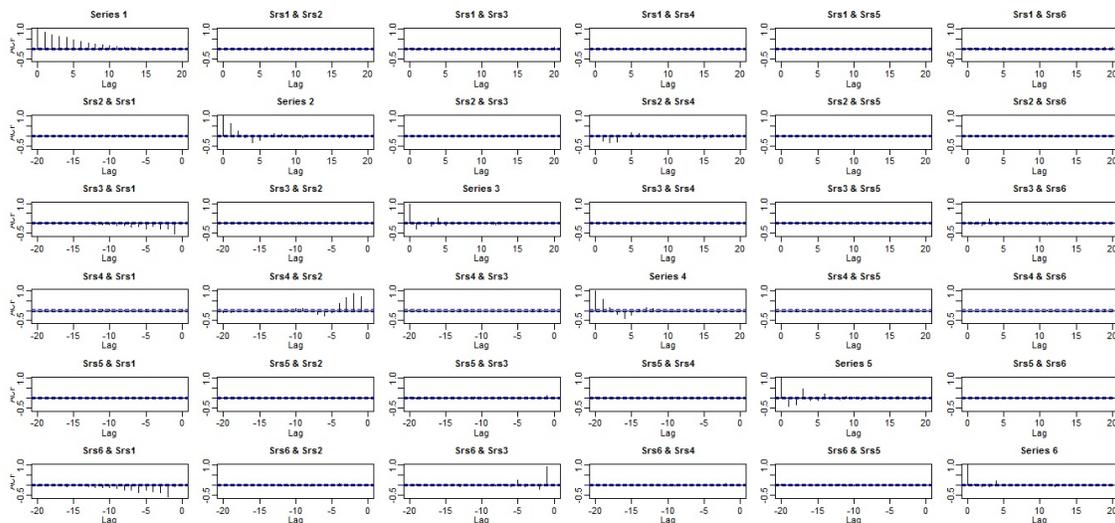


Figure 13: Sample moving cross-correlation plots with window size 101 for transformed data after GTS-PCA is applied.

Moreover, by replicating the above simulation 500 times, GTS-PCA with  $W = 101$  was able to extract the correct segmentations 97% of the time. This is the same percentage that was reported reported by Chang et al. (2018) when TS-PCA was used. This indicates that both methods have identical performance on stationary series, even if GTS-PCA uses a window that is much smaller than the ideal size.

By the end of the above simulation studies, we conclude that GTS-PCA was able to reduce the dimension of a multivariate time series into uncorrelated subseries with lower dimensions. Our simulations showed that GTS-PCA performed well when applied to series with large sample sizes.

## 5.2 Real Data Examples

**Example 5.** In this example, we apply GTS-PCA on a non-stationary series that consists of the daily stock prices of six US companies in US Dollar from Nov 07, 2013 to Dec 18, 2017. The total number of observations is 1036 days. The six companies are CVS health (CVS), Chevron (CVX), Express Scripts Holding (ESRX), Intel (INTC), Lowe's (LOW) and Prudential Financial (PRU). The data was obtained from *Yahoo!* Finance.

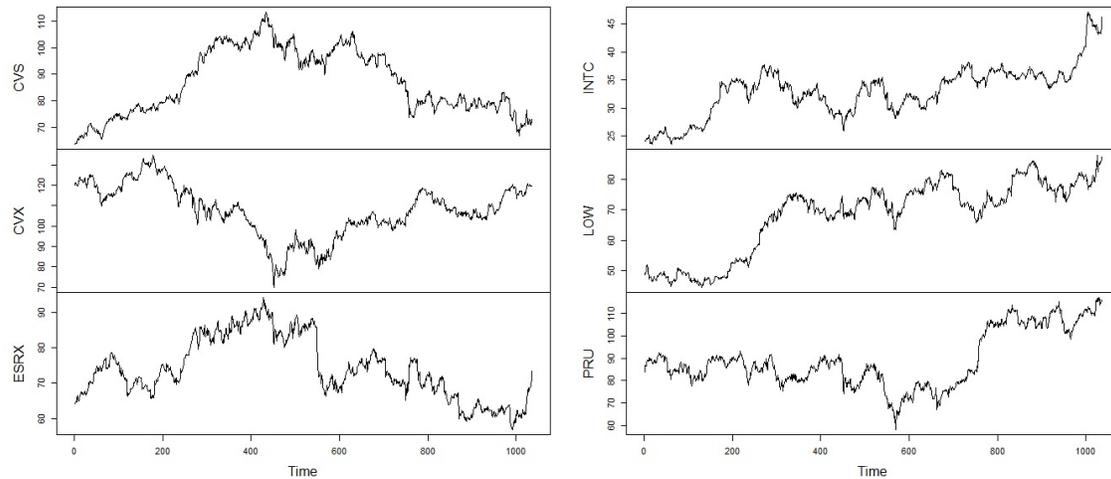


Figure 14: A time series plot of the daily stock prices of the 6 US companies

The six companies showed non-stationary behaviours over time, according to the time series plot in Figure 14. GTS-PCA with  $W = 101$  and  $l_1 = 5$  is used to reduce the dimension of the series. The sample moving cross-correlation plots with  $W = 101$  between the six companies before applying GTS-PCA is provided in Figure 15. A moderated dynamic relationship existed between the six companies. For example, the company INTC had a moderate relationship with LOW. The companies INTC and PRU had a strong relationship with CVX for negative lagged values. Moreover, the company ESRX was weakly correlated with INTC.

After GTS-PCA was applied, the moving cross-correlation with  $W = 101$  suggested dividing the data into four subseries with three univariate subseries and one multivariate subseries that consisted of the companies CVS, ESRX and LOW; See Figure 16.

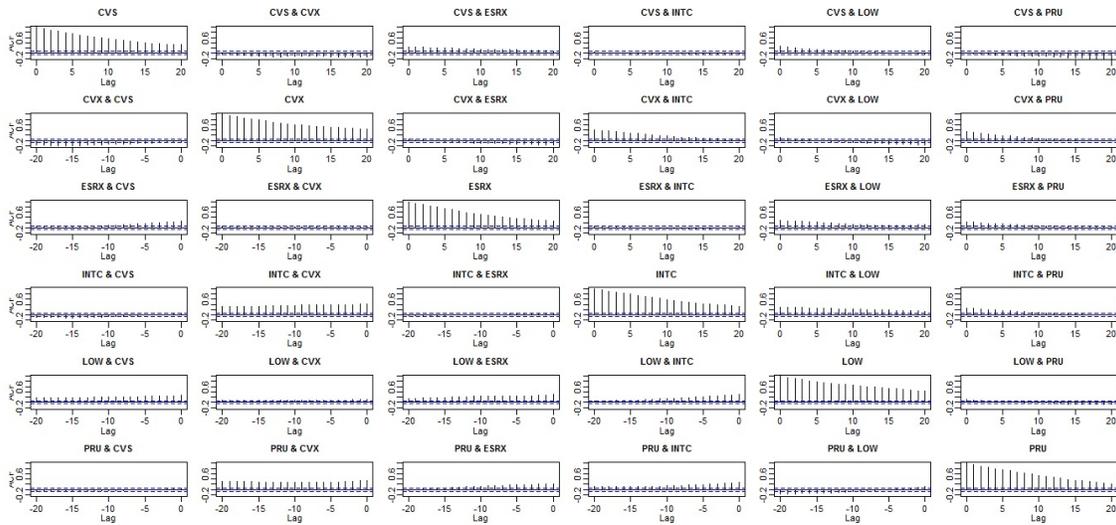


Figure 15: Sample moving cross-correlation plots of the daily stock prices of the 6 US companies before GTS-PCA is applied.

Therefore, we conclude that GTS-PCA was able to reduce the dimension of  $\mathbf{z}_t$  from six to four uncorrelated series of dimensions 3, 1, 1 and 1. These four subseries can be analysed separately as they are uncorrelated statically and dynamically.

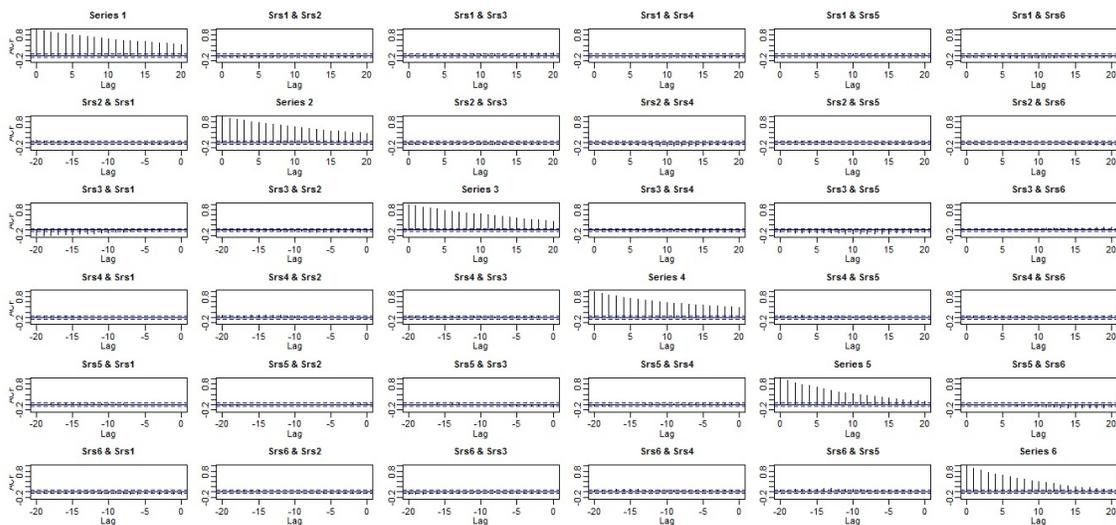


Figure 16: Sample moving cross-correlation plots of the daily stock prices of the 6 US companies after the GTS-PCA is applied.

**Example 6.** In the following example, we apply the GTS-PCA on a non-stationary data

set with 22 variables. The data set in this example consists of the daily stock prices of 22 US companies in US Dollar from Nov 07, 2013 to Dec 18, 2017 by *Yahoo!* Finance. The total number of observations is 1036 days. The companies' names and labels are listed in Table 5.

No.	Label	Name	No.	Label	Name
1	CVS	CVS Health	12	KR	Kroger-The Kroger
2	CVX	Chevron	13	LOW	Lowe's
3	ESRX	Express Scripts Holding	14	MCK	McKesson
4	FB	Facebook	15	MET	MetLife Insurance
5	F	Ford Motor	16	MSFT	Microsoft
6	GE	General Electric	17	PEP	PepsiCo
7	GM	General Motor	18	PG	The Procter & Gamble
8	HD	The Home Depot	19	PRU	Prudential Financial
9	IBM	IBM	20	PSX	Philips 66
10	INTC	Intel	21	TGT	Target
11	JNJ	Johnson & Johnson	22	TWTR	Twitter

Table 5: Names and labels of the 22 US companies.

A time series plot of the daily stock prices of the 22 companies is shown in figures 17, 18 and 19. The non-stationary behaviour of the variables can be clearly seen from the time series plot.

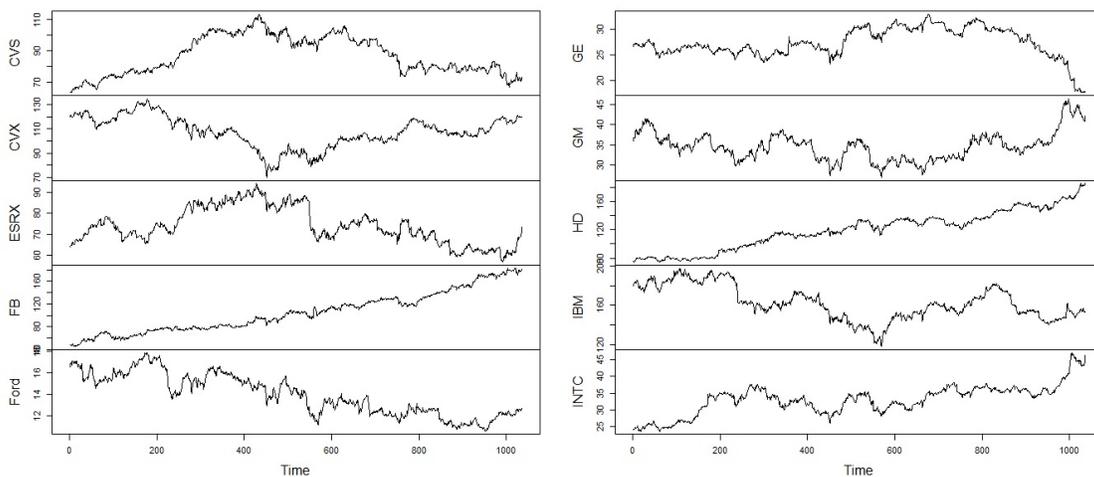


Figure 17: A time series plot of the daily stock prices of the 22 US companies - Part 1

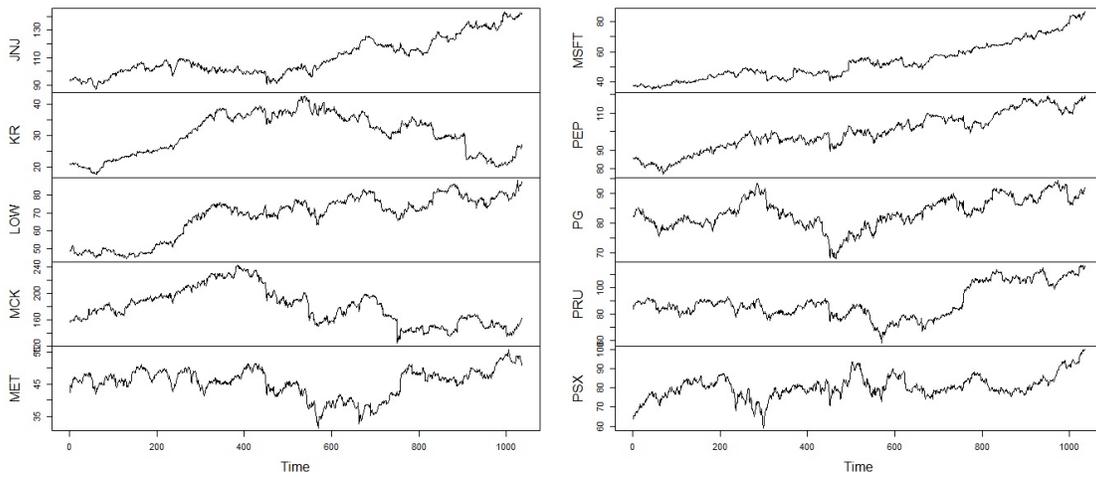


Figure 18: Time series plot of the daily stock prices of the 22 US companies - Part 2

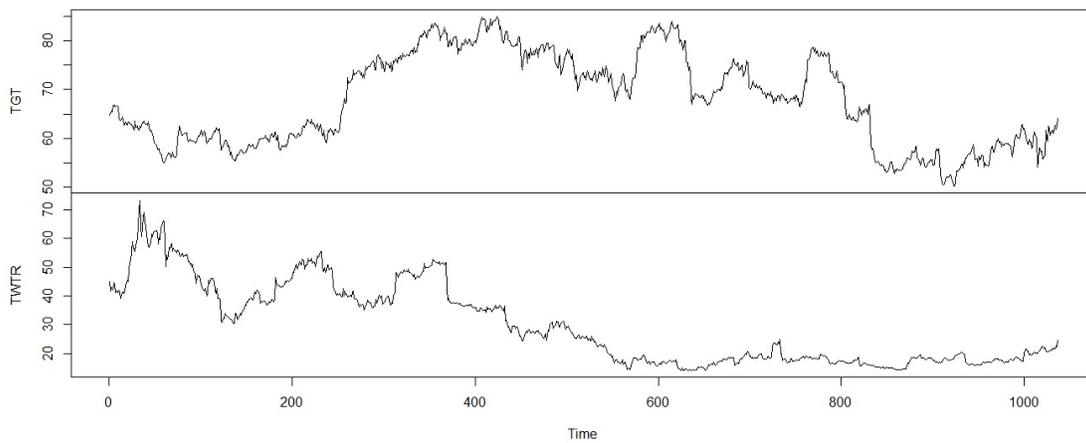


Figure 19: Time series plot of the daily stock prices of the 22 US companies - Part 3

Figures 20, 21, 22 and 23 show only the first and last 50 sample moving cross-correlation plots with  $W = 101$  of the 22 companies before applying GTS-PCA, where a moderated relationship existed between the 22 companies.

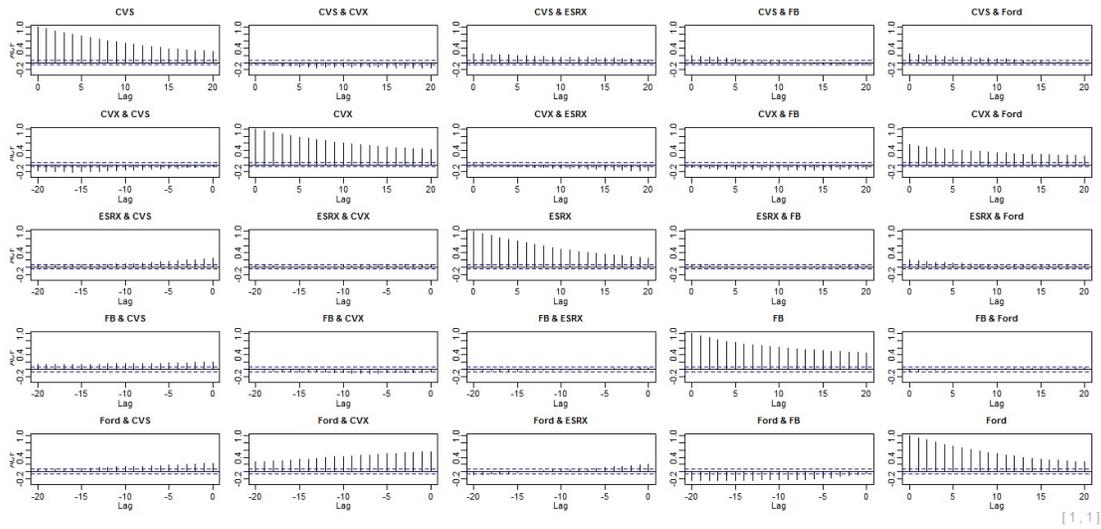


Figure 20: First 50 sample moving cross-correlation plots of the daily stock prices of the 22 US companies before GTS-PCA is applied - Part 1.

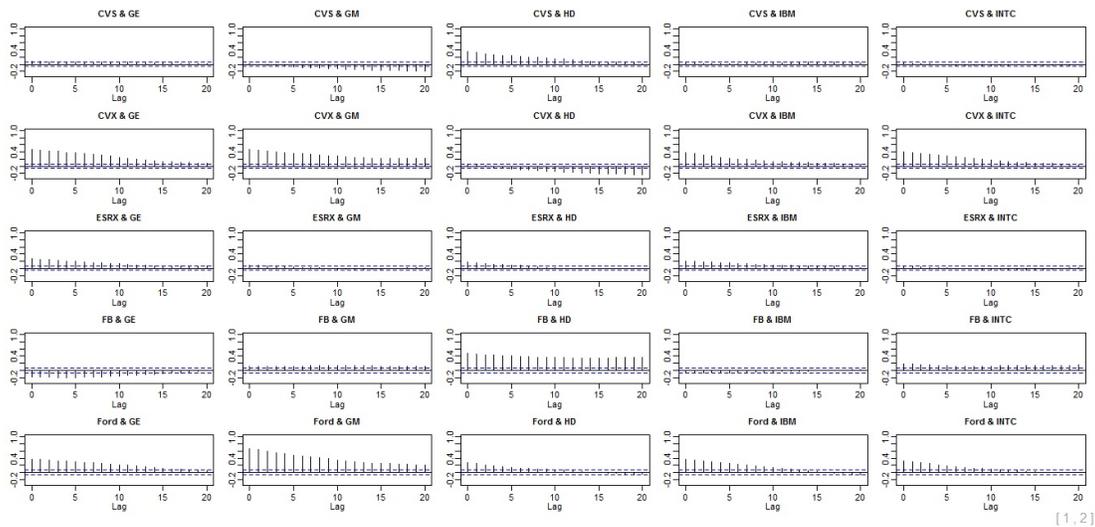


Figure 21: First 50 sample moving cross-correlation plots of the daily stock prices of the 22 US companies before GTS-PCA is applied - Part 2.

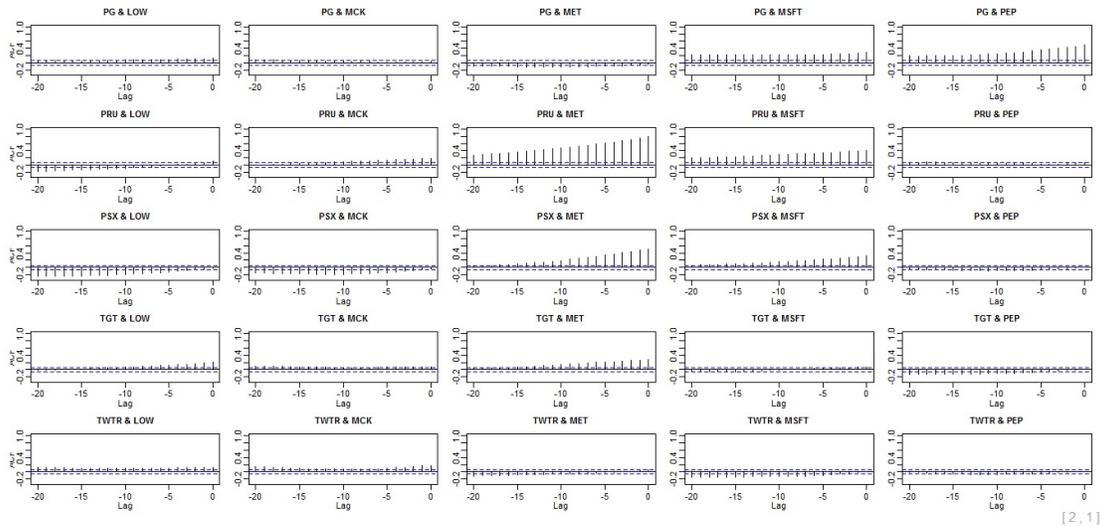


Figure 22: Last 50 sample moving cross-correlation plots of the daily stock prices of the 22 US companies before GTS-PCA is applied - Part 1.

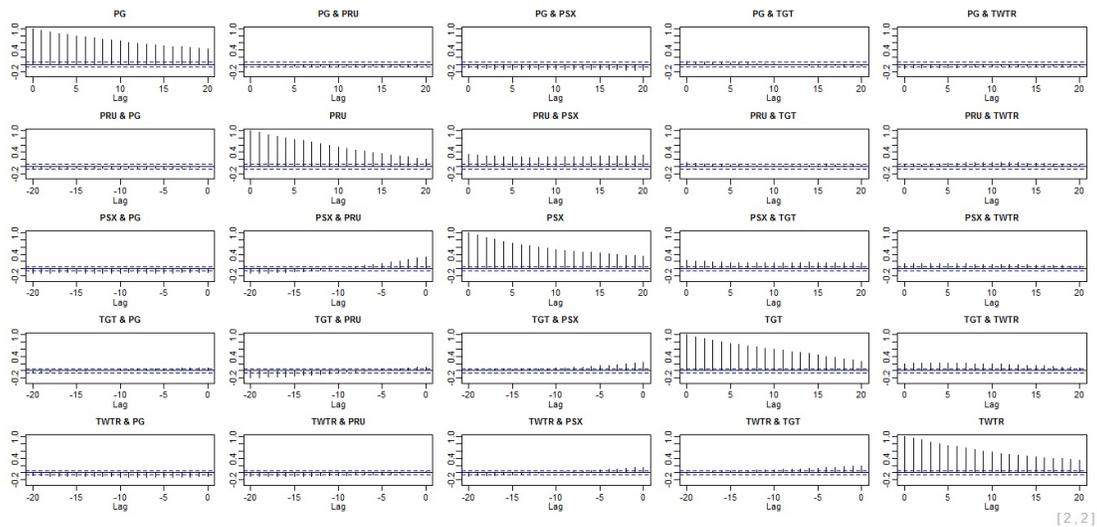


Figure 23: Last 50 sample moving cross-correlation plots of the daily stock prices of the 22 US companies before GTS-PCA is applied - Part 2.

After applying GTS-PCA with  $W = 101$  and  $l_1 = 5$  to the daily stock prices, the maximum moving cross-correlation with  $W = 101$  suggested to segment the data into 15 subseries that consisted of 11 univariate subseries, three bivariate subseries and one multivariate time series with five variables, as shown in Table 6. Further analyses can be applied separately to these 15 subseries since they are uncorrelated statically and dynamically.

Results Method	Correlated Variables after applying GTS-PCA with (W=101)
Maximum Moving Cross-Correlation (W=101)	{CVX, F, HD, KR, MET}, {ESRX, LOW}, {GE, MSFT} and {JNJ, MCK}

Table 6: Summary results of the transformed segmentation of the 22 US companies after applying GTS-PCA with  $W = 101$

So we were able to reduce the dimension of the non-stationary daily stock prices of 22 US companies to 15 uncorrelated subseries by using GTS-PCA with a window size of 101 and 5 lagged series.

## 6 Concluding Remarks

In summary, GTS-PCA is a PCA-based dimension reduction method that can reduce the dimension of both stationary and non-stationary time series to uncorrelated subseries, where further analyses can be done separately on these subseries. TS-PCA can be considered as a special case of GTS-PCA where  $W = T$ , and they produce the same segmentation on stationary data; See Example 4.

GTS-PCA transform data by applying eigenanalysis on  $\hat{\mathbf{M}}\mathbf{W}_z$  defined in (2.12). The moving cross-correlation function of Alshammri and Pan (2019) can be used to rearrange the components of the transformed series with moderate dimensions into subseries based on their dynamic relationship. The maximum moving cross-correlation is a new proposed tool that is based on the moving cross-correlation function and can be used to rearrange the components of transformed series with higher dimensions into correlated subseries.

Choosing the window size for GTS-PCA depends on the stationarity of the data. GTS-PCA with smaller window sizes is more suitable for data with stronger non-stationarity, and the opposite is true. Based on simulated data, we noticed that GTS-PCA was able to produce accurate results even when it used window sizes that are smaller than required. A further research idea is to develop a criteria that can objectively determine the optimal window size for GTS-PCA.

We studied the convergence of the estimator  $\hat{\mathbf{A}}$  when the dimension  $m$  is fixed and when  $m$  depends on the window size  $W$  for large dimensional cases. We stated the needed

assumptions to obtain the convergence rates for these cases and showed that  $\hat{\mathbf{A}}$  is a consistent estimator of  $\mathbf{A}$ . For ultra high dimensions, where  $m$  depends on  $W$ , a threshold estimator of the sample cross-covariance matrix over each window  $\hat{\Gamma}_{z,i}(l)$  should be used to preserve the convergence of  $\hat{\mathbf{A}}$ .

Our simulations showed that GTS-PCA performs reasonably on time series with different dimensions and sample sizes. We also tested GTS-PCA on real data sets. A drawback of GTS-PCA is that the transformed uncorrelated subseries might not exist. This might occur in some cases where all the transformed series are correlated. A similar conclusion was reported for the stationary TS-PCA; See Chang et al. (2018).

## DATA AVAILABILITY STATEMENT

The data used to support the findings of this article are available at Yahoo! Finance via <https://finance.yahoo.com>.

## REFERENCES

- Alshammri, F. (2020a). *GTSPCA: Generalized Principal Component Analysis for Non-Stationary Vector Time Series*. R package version 0.1.0.
- Alshammri, F. (2020b). *MpermutMax: The Maximum Moving Cross-Correlation Method*. R package version 0.1.0.
- Alshammri, F. and Pan, J. (2019). Moving dynamic principal component analysis for non-stationary multivariate time series. Manuscript submitted for publication.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.
- Box, G. E. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika*, 64(2):355–365.
- Brillinger, D. (1981). *Time Series: Data Analysis and Theory*, volume 36. Classics in Applied Mathematics, SIAM.

- 
- Chang, J., Guo, B., and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, 46(5):2094–2124.
- Chen, X., Xu, M., and Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251–276.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840.
- Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- Ku, W., Storer, R. H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1):179–196.
- Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.
- Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.
- Peña, D. and Box, G. E. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82(399):836–843.
- Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136(4):1237–1257.
- Peña, D. and Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, 111(515):1121–1131.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83(404):1097–1107.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–213.

Zhang, R., Robinson, P., and Yao, Q. (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association*, 114(526):916–927.

## APPENDIX

### Proofs: Asymptotic Results ( $m$ is fixed)

In order to prove Theorem 1, we present a few lemmas first.

**Lemma 1.** *Assume the dimension  $m$  is fixed. Under assumptions 1 and 2, for each  $i$ , it holds that*

$$\|\hat{\Gamma}_{z,i}(l) - \Gamma_{z,i}(l)\|_2 = O_p(1/\sqrt{W}),$$

and therefore

$$\|\hat{\Gamma}_{z,i}(l) - \Gamma_{z,i}(l)\|_2 \xrightarrow{P} 0$$

for all  $l \leq l_1$  as  $W \rightarrow \infty$ , where  $\Gamma_{z,i}(l)$  and  $\hat{\Gamma}_{z,i}(l)$  are defined in (2.2) and (2.3), respectively.

This lemma can be obtained in a similar way to the proof of Lemma 2 in the supplementary materials of Chang et al. (2018) by noting that the condition in assumption 1 is assumed uniformly over  $i$ .

**Lemma 2.** *Under assumptions 1-3,*

$$\|\hat{\mathbf{M}}\Gamma_z(l) - \mathbf{M}\Gamma_z(l)\|_2 \xrightarrow{P} 0$$

for all  $l \leq l_1$  as  $W \rightarrow \infty$ , where  $\mathbf{M}\Gamma_z(l)$  and  $\hat{\mathbf{M}}\Gamma_z(l)$  are defined in (2.4) and (2.5), respectively.

**Proof of Lemma 2.** From the definitions of  $\mathbf{M}\Gamma_z(l)$  and  $\hat{\mathbf{M}}\Gamma_z(l)$ , and by assuming 1-3 hold and  $m$  is fixed, then

$$\begin{aligned} \|\hat{\mathbf{M}}\Gamma_z(l) - \mathbf{M}\Gamma_z(l)\|_2 &= \left\| \frac{1}{T-2w} \sum_{i=w+1}^{T-w} \hat{\Gamma}_{z,i}(l) - \frac{1}{T-2w} \sum_{i=w+1}^{T-w} \Gamma_{z,i}(l) \right\|_2 \\ &\leq \frac{1}{T-2w} \sum_{i=w+1}^{T-w} \|\hat{\Gamma}_{z,i}(l) - \Gamma_{z,i}(l)\|_2 \\ &\xrightarrow{P} 0. \end{aligned}$$

**Lemma 3.** Under assumptions 1 and 2, then

$$\|\hat{\mathbf{M}}\mathbf{W}_z - \mathbf{M}\mathbf{W}_z\|_2 \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ .

**Proof of Lemma 3.** From Lemma 2 above, we know that

$$\|\hat{\mathbf{M}}\Gamma_z(l) - \mathbf{M}\Gamma_z(l)\|_2 \xrightarrow{P} 0$$

for all  $l \leq l_1$ , and by using the definition of  $\mathbf{M}\mathbf{W}_z$  in (2.10), then the result follows immediately.

Now, we are ready to prove Theorem 1.

**Proof of Theorem 1.** Let assumptions 1-3 hold and the dimension  $m$  is fixed. Applying the results of Theorem 8.1.10 of Golub and Van Loan (1996), we have

$$\max_{1 \leq j \leq n} D(\mathcal{M}(\hat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j)) = O_p(\|\hat{\mathbf{M}}\mathbf{W}_z - \mathbf{M}\mathbf{W}_z\|_2)$$

Then, from the results of Lemma 3 above, we get

$$\max_{1 \leq j \leq n} D(\mathcal{M}(\hat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j)) \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ .

### Proofs: Asymptotic Results ( $m = o(W^c)$ )

We are going to prove Theorem 2 based on a few lemmas. Recall, the estimators in (4.4), (4.5) and (4.6) are used instead of the ones in (2.3), (2.5) and (2.10), respectively, where

the threshold level is set to  $u = cm^{2/\gamma} W^{-1/2}$ . The following lemma can be proved by the similar way to that in Chang et al. (2018).

**Lemma 4.** *For an  $m$ -dimensional series  $\mathbf{z}_t$  where  $m = o(W^{\gamma/4})$ , under assumptions 4-6, for each window  $\mathbf{z}_i$ , we have*

$$\|T_u(\hat{\Gamma}_{\mathbf{z},i}(l)) - \Gamma_{\mathbf{z},i}(l)\|_2 = O_p((m^{4/\gamma}W^{-1})^{(1-\delta)/2}f),$$

where  $f$  is defined in (4.11).

Therefore we have the following lemma based on the result of Lemma 4 above.

**Lemma 5** *Let assumptions 4-6 hold. Assume that  $m = o(W^{\gamma/4})$ . Then we have*

$$\|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2 \xrightarrow{P} 0$$

for all  $l \leq l_1$ , as  $W \rightarrow \infty$ .

**Proof of Lemma 5** Under assumptions 4-6, applying Lemma 4 to each window  $\mathbf{z}_i$ , we get

$$\|T_u(\hat{\Gamma}_{\mathbf{z},i}(l)) - \Gamma_{\mathbf{z},i}(l)\|_2 \xrightarrow{P} 0$$

for all  $l \leq l_1$  as  $W \rightarrow \infty$ , provided  $m = o(W^{\gamma/4})$ . By following similar argument to the proof of Lemma 2, the proof of this lemma is complete.

**Lemma 6.** *Let the assumptions 4-6 hold. Assume further that  $m = o(W^{\gamma/4})$ . Then it holds that, as  $W \rightarrow \infty$ ,*

$$\|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2 \xrightarrow{P} 0.$$

**Proof of Lemma 6.** Since

$$\begin{aligned} \|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2 &= \|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) [\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l)]' - \mathbf{M}\Gamma_z(l) \mathbf{M}\Gamma_z'(l)\|_2 \\ &\leq 2\|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2 \|\mathbf{M}\Gamma_z(l)\|_2 + \|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2^2 \\ &\leq 2\|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2 \|\mathbf{M}\Gamma_x(l)\|_2 + \|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2^2. \end{aligned}$$

Recall that  $h = \max_{1 \leq l \leq l_1} \|\mathbf{M}\Gamma_x(l)\|_2$ , then by Lemma 6, we have

$$\|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2 \xrightarrow{P} 0.$$

Now, we are ready to prove Theorem 2.

**Proof of Theorem 2.** Under assumptions 4-7. Assume further that  $m = o(W^{\gamma/4})$ .

Then by applying the results of Theorem 8.1.10 of Golub and Van Loan (1996), we have

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) = O_p(\|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2).$$

By Lemma 6, we get

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) \xrightarrow{P} 0.$$

### Proofs: Asymptotic Results ( $\log m = o(W^c)$ )

The following lemmas are needed to prove Theorem 3. Recall that we are using the estimators in (4.4), (4.5) and (4.6) but with the threshold level  $u = c(W^{-1} \log m)^{1/2}$ .

**Lemma 7.** *Let assumptions 4, 8 and 9 hold. Let  $\log m = o(W^{b/(2-b)})$  where  $b = 1/(2\eta^{-1} + \xi^{-1})$ . Then*

$$\|\hat{\mathbf{M}}\Gamma_z^{\text{thre}}(l) - \mathbf{M}\Gamma_z(l)\|_2 \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ .

**Proof of Lemma 7.** Notice that  $\log m = o(W^{b/(2-b)})$  and  $u = c(W^{-1} \log m)^{1/2}$ . Under assumptions 4, 8 and 9, if we apply Lemma 4 on each window  $\mathbf{z}_i$ , then it follows

$$\|T_u(\hat{\Gamma}_{z,i}(l)) - \Gamma_{z,i}(l)\|_2 = O_p(f(W^{-1} \log m)^{(1-\delta)/2}).$$

Therefore, noting that  $W^{-1} \log m \rightarrow 0$ , we have, for each  $i$ ,

$$\|T_u(\hat{\Gamma}_{z,i}(l)) - \Gamma_{z,i}(l)\|_2 \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ . The rest of the proof can be obtained by following similar argument to the proof of Lemma 2.

**Lemma 8.** *Let assumptions 4, 8 and 9 hold. Then for  $\log m = o(W^{b/(2-b)})$ , it follows*

$$\|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2 \xrightarrow{P} 0$$

as  $W \rightarrow \infty$ , where  $b = 1/(2\eta^{-1} + \xi^{-1})$ .

**Proof of Lemma 8.** Using the results from Lemma 7, the proof of this lemma can be obtained by following similar steps to the proof of Lemma 6.

**Proof of Theorem 3.** Under assumptions 4, 7, 8 and 9. Let  $\log m = o(W^{b/(2-b)})$  where  $b = 1/(2\eta^{-1} + \xi^{-1})$ . Theorem 8.1.10 of Golub and Van Loan (1996) yields

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) = O_p(\|\hat{\mathbf{M}}\mathbf{W}_z^{\text{thre}} - \mathbf{M}\mathbf{W}_z\|_2).$$

Then, by using Lemma 8, we complete the proof as

$$\max_{1 \leq k \leq n} d_k D(\mathcal{M}(\hat{\mathbf{A}}_k), \mathcal{M}(\mathbf{A}_k)) \xrightarrow{P} 0.$$