Dimension Reduction of Stationary Multivariate Time Series

Researcher in Statistics at the University of Strathclyde, UK

Fayed Alshammri

Introduction

Chang et al. (2016) extended PCA by finding a linear transformation of the original variables such that the transformed series is segmented into uncorrelated subseries with lower dimensions. This method is called TS-PCA. In my current research, I will extend TS-PCA by reducing the dimension of the transformed subseries further by applying GDPCA by Pena and Yohai (2016) to the results from TS-PCA, and possibly reach a further dimension reduction. Hence, the proposed method is a combination of TS-PCA and GDPCA.

Methods

Let the vector $\boldsymbol{Z}_t$ be an m-dimensional time series. To reach further dimension reduction, use the following:

TS-PCA:

- Find a linear transformation $\boldsymbol{Z}_t = \mathbf{A} * \mathbf{X}_t$ such that $\mathbf{X}_t$ is segmented into q uncorrelated subseries $\mathbf{X}_{i,t}$, where $1 \leq i \leq q$ and q<m.
- **A**, can be found by applying Eigen-analysis to a quadratic function of the auto-covariance matrix of standardized $\boldsymbol{Z}_t$ up to some lag l_1.

GDPCA:

- For each multivariate subseries $\mathbf{X}_{i,t}$ from TS-PCA results, set $\mathbf{Y}_t = \mathbf{X}_{i,t}$.
- Reconstruct $\mathbf{Y}_t$ such that $\mathrm{MSE}(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{1}{Tm}\sum_{j=1}^{m}\sum_{t=1}^{T}(Y_{j,t} - \sum_{i=0}^{k}\beta_{j,i+1} f_{t+i} - \alpha_j)^2$ is minimized.
- $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be estimated by Least Squared.
- An iteration algorithm is used to enhance the results.

http://redislabs.bluecoastweb.com

Data Analysis

A multivariate time series $\boldsymbol{Z}_t$ with 6 variables and 2000 observations is simulated in the software R,

We apply the proposed method to reach further dimension reduction. Data analysis is summarized as follows:

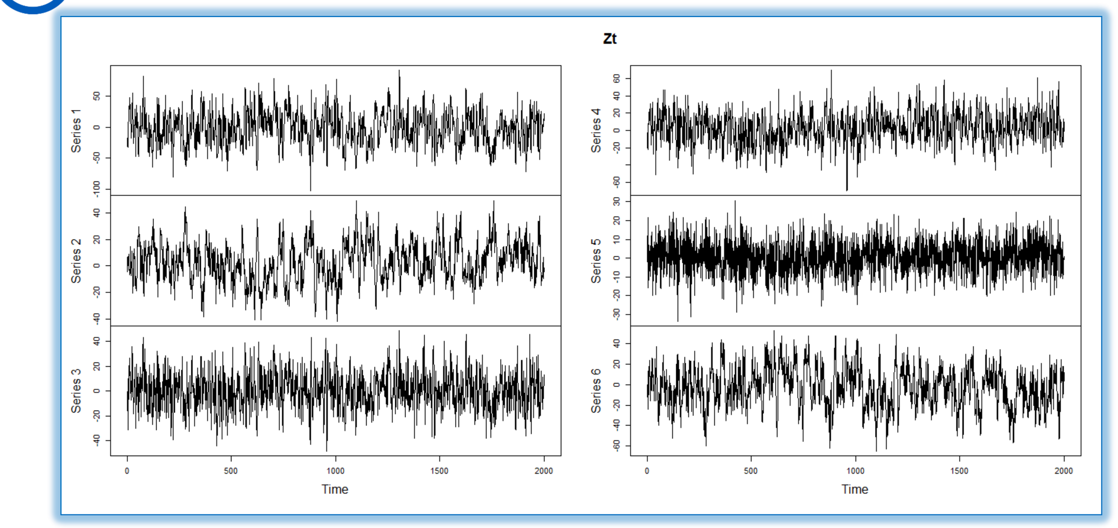
A. The time series plot of the simulated data $\boldsymbol{Z}_t$ is shown in Figure A.

B. Apply TS-PCA to $\boldsymbol{Z}_t$ in order to find the hidden uncorrelated subseries $\mathbf{X}_{i,t}$. The results are shown in Figure B.

C. Apply GDPCA to each multivariate subseries from step B separately. We use l_1 equal to 10 lags and 200 iterations. See Figure C.

A

Time Series Plot



B

ACF Plot

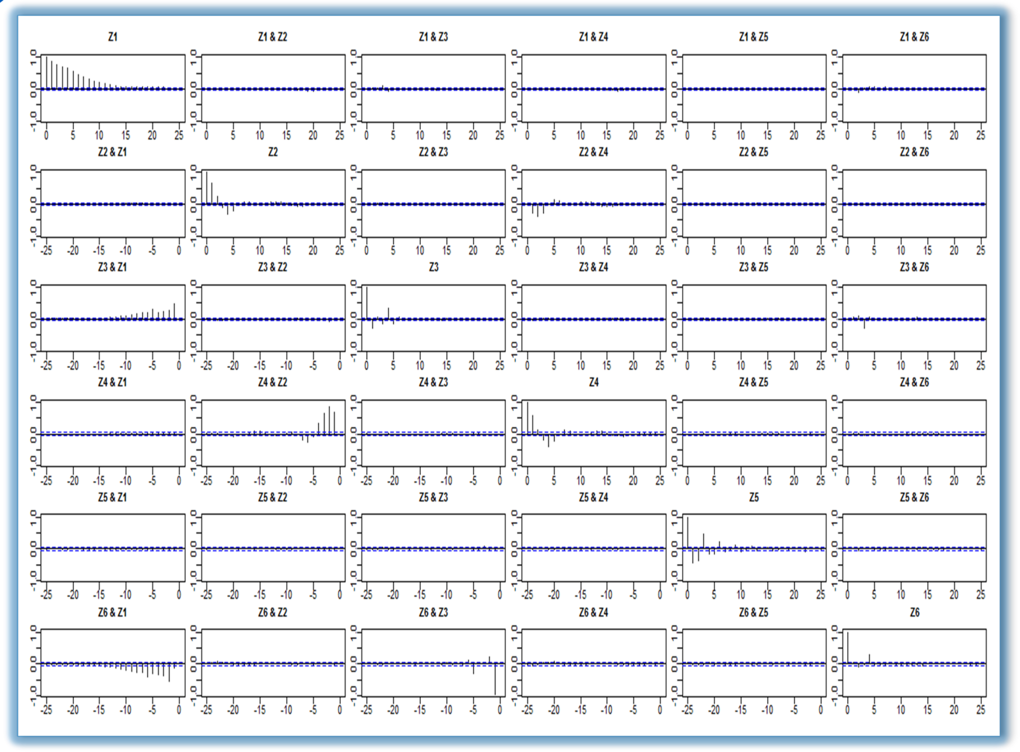


Figure B. A plot of the autocorrelation function of the transformed series X_t.

C

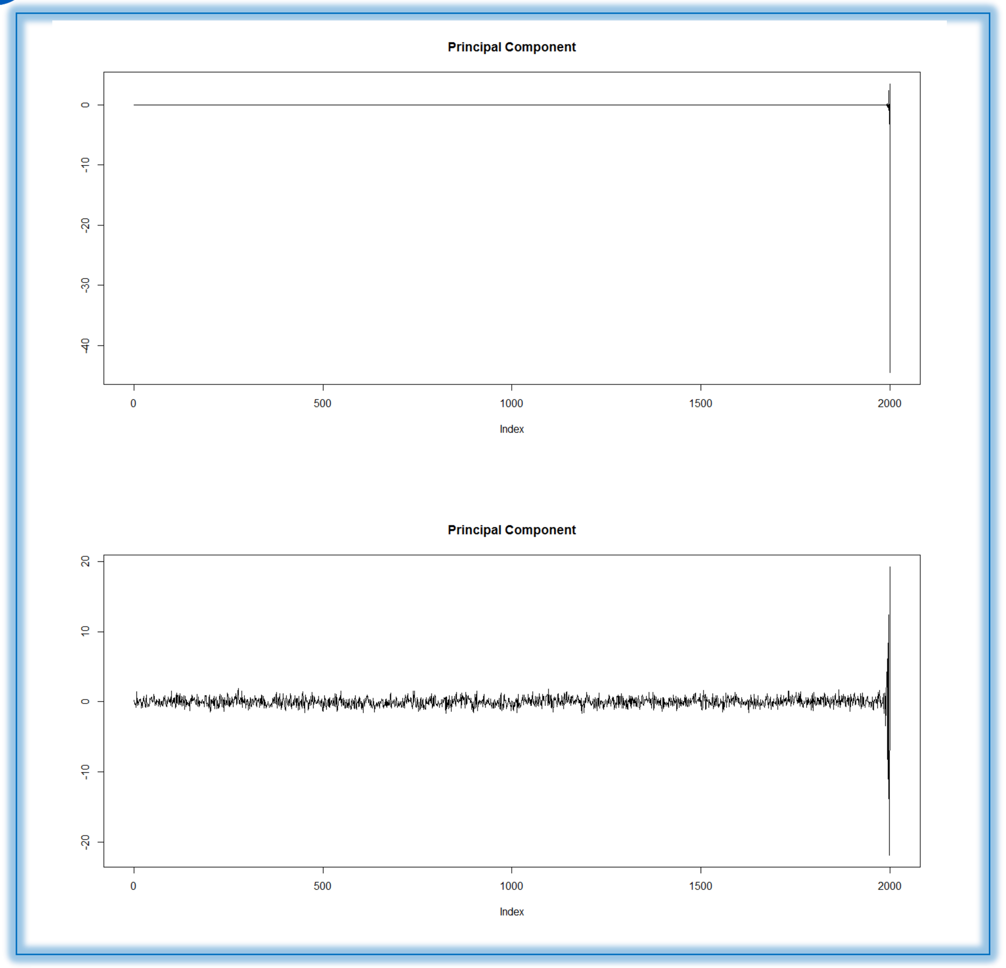
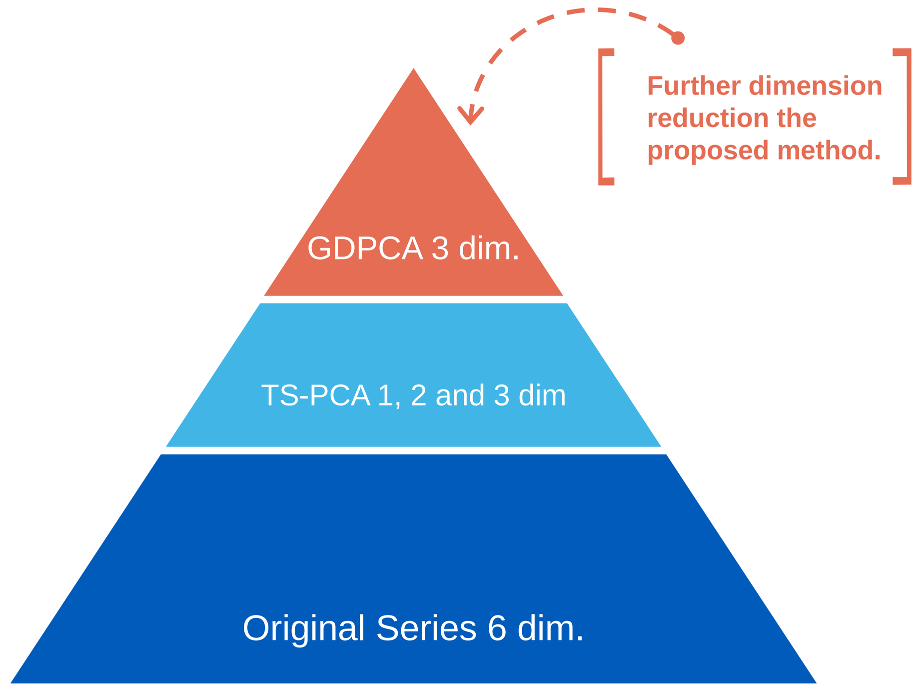


Figure C. Plot of the Dynamic Component used to reconstruct $\mathbf{X}_{2,t}$ (on top) and $X_{3,t}$ (on bottom).

Results

- A TS-PCA reduces the dimension of $\mathbf{Z}_t$ by segmenting it into 3 uncorrelated subseries $\mathbf{X}_{1,t}$, $\mathbf{X}_{2,t}$ and $\mathbf{X}_{3,t}$ with dimensions 1, 2 and 3 respectively.
- By Applying GDPCA to the series $\mathbf{X}_{2,t}$, we reduced its dimension from 2 to 1 with 0.1% MSE and 99% of the variability explained. Similarly, by applying GPDCA to the series $\mathbf{X}_{3,t}$ we reduce its dimension from 3 to 1 0.2% MSE and 98% of the variability explained.



Conclusion

By combining the dimension reduction methods, namely TS-PCA and DGPCA, we were able to reduce the dimension of the multivariate time series $\boldsymbol{Z}_t$ from 6 to only 3 dimensions, where each one can be analyzed separately as they are uncorrelated series. Therefore, we conclude with the following:

- **Use TS-PCA first:** An advantage of TS-PCA is its ability to segment the data into subgroups that can be analyzed separately.
- **Use GDPCA next:** To reduce the dimension of the segments from TS-PCA even further.
- **An important reason to choose the above order:** Forecasting using TS-PCA segments is much more accurate than other methods, hence, we would like to have as much uncorrelated univariate transformed subseries as possible from TS-PCA, then finally use GDPCA to reduce the dimension of the multivariate once. In the other hand, if forecasting is not one of your goals, then GDPCA would be the best choice because it has the ability to reconstruct a data with a large number of variables using only one components with a very low percentage of error and very hay percentage of variability explained.

References

- Chang, J., Guo, B. and Yao, Q. (2016). Segmenting multiple time series by contemporaneous linear transformation: PCA for Time Series. Available at arXiv:1410.2323v3.
- Pena, D. and Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, 111(515), 1121-1131.
- Tsay, R. (2014). *Multivariate Time Series Analysis*. Wiley.