Title: A Method of Attaining Sparsity of Principal Components in Time Series Data

Author: Joseph Ryan G. Lansangan

Degree: Master of Science in Statistics

Date: April 2006

Abstract:

Dimension reduction have been extensively used for cross-sectional and time-series data. One of the common tools for dimension reduction is Principal Component Analysis (PCA). A possible drawback however is the difficulty in interpretation of the first few principal components. Approaches such as rotation and thresholding are meant to address the interpretation problem. Rotation is not recommended for temporal observations, but thresholding may yield misleading results. Sparse Principal Component Analysis (SPCA) is another option proposed by (Zou et. Al.,2004), where the loadings of unimportant variables are set to zero.

This study investigates a method of attaining sparsity that seldom occurs when the variables that are subjected to PCA are non-stationary. It focuses on exploring PCA and SPCA on non-stationary AR(1) time series data characterized by a drift-in-mean over time where the number of time series exceeds the number of observations (timepoints). A mathematical characterization of PCA explains why and how interpretation of the first few components becomes complicated. Simulations and application to an agricultural data were also made. The empirical results identified the parameter settings that produce convergence of the algorithm under different scenarios. The cross-correlations (lag 0) are important determinants of the results of SPCA.

The results show that sparsity can be attained in constructing principal components of non-stationary (drift in mean) time series. SPCA may give non overlapping loadings without losing a large amount of explained variability; hence ease in interpretation of the components at acceptable level of predicted variability. Intervals of the iteration parameters are suggested to obtain convergence of the estimation procedure.