2 - Principal Components Analysis

(a) Introduction



Principal components analysis (PCA) is a multivariate technique that analyzes a data in which observations are described by several inter-correlated quantitative dependent variables. This technique extracts the important information from the data, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations.

Factor models that offer explanations of stock returns and correlations have been very popular in finance. PCA is unlike traditional factor models because the factors it creates do not usually have an economic interpretation. Rather, the components(factors) constructed in PCA are built to have special statistical characteristics:

• Each component accounts for as much variation in the underlying data as possible.

• Each component is uncorrelated with every other factor.

• Principal components elucidate the dominant combinations of variables within the covariance structure of the data.

Mathematically, pca depends upon the eigen-decomposition of positive semi-definite matrices and upon the singular value decomposition (svd) of rectangular matrices.

(b) Mathematics of Principal Components

Singular Value Decomposition

Properties of a data matrix – first and second moments

Principal component analysis (PCA) by SVD

(C) Implementation