Reduced-Rank Regression Model: A Review

Frances Lin

March 2022

Abstract

Background and Introduction

A classical multivariate linear model, which is given as

$$Y_k = CX_k + \varepsilon_k, \; k = 1, ..., T$$

where $Y_i = (y_{1k}, ...y_{mk})^T$ is a $mx1$ response vector, C is a mxn regression coefficient matrix, $X_k = (x_{1k}, ...x_{mk})^T$ is a $nx1$ predictor vector, and $\varepsilon_k = (\varepsilon_{1k}, ...\varepsilon_{mk})^T$ is a $mx1$ error vector with $E(\varepsilon_k) = 0$ and $cov(\varepsilon_k) = \Sigma_{\varepsilon\varepsilon}$, does not make use of the fact that the response variables are likely correlated. In addition, in many practical situations, there is often a need to reduce the number of parameters in the model since it can be too large.

Further assuming reduced rank of the matrix C such that

$$rank(C) = r \leq min(m, n),$$

leads us to two implications. First, let l be the constraint vector, the linear combination, $l^T Y_k$, $i = 1, ..., (m-r)$, can be modeled through the distribution of the error term ε_k (w/o X_k). Second, C can be expressed as $C = AB$, where A is of dimension mxr and B is of dimension rxn. Then, the above multivariate linear model can be rewritten as

$$Y_k = A(BX_k) + \varepsilon_k, \; k = 1, ..., T,$$

where BX_k is of reduced dimension $rx1$, and as a result, there is a gain in simplicity and interpretation since the r linear combinations of the predictors X_k are sufficient to model the response variables Y_k.

The first application of reduced-rank regression model appeared in an initial work of Anderson (1951) in the field of economics. The model and its statistical properties were further examined by a few other authors. Subsequent but separate work that were studied using related concepts were principle components (Rao, 1964), simultaneous linear prediction modeling (Fortier, 1966), redundancy analysis, an alternative to canonical correlation analysis (van den Wollenberg, 1977), etc. More complex models have also been developed ever since.

Applications

Applications of the reduced-rank regression model include (1) the experimental properties of hydrocarbon fuel mixtures in relating response to composition (Davies and Tso, 1982), (2) an econometric model of the United Kingdom from 1948 to 1956 (Gudmundsson, 1977), which consists of 37 time series of response variables and 32 time series of predictors, (3) the relationship between measurements on solar radiation taken over various sites in Scotland and the physical characteristics of the sites (Glasbey, 1992), (4) the joint effects of toxic compounds on the growth of larval fathead minnows (Ryan et al., 1992), and (5) testing the efficiency of portfolios (Zhou, 1991, 1995).

Estimation

The matrix A of dimension mxr, B of dimension rxn and $\Sigma_{\varepsilon\varepsilon}$ of the error term are the parameters that are to be estimated.

Presentation of reduced-rank estimation requires an important matrix result called the Householder-Young or Eckart-Young Theorem (Eckart and Young, 1936).

Discussion

Reference