Reduced-Rank Regression Model: A Review

Frances Lin

March 2022

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 (or imposing) 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

$$Y_k = ABX_k + \varepsilon_k,\ k = 1, ..., T,$$

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

Reduced-rank (RR) estimation is obtained as a certain RR approximation of the full-rank least squares estimate of the coefficient matrix. Therefore, to present the estimation, we need the Brillinger's theorem (Brillinger, 1981, Section 10.2), which can be proven by the Eckart-Young theorem (Eckart and Young, 1936).

Brillinger's Theorem. Suppose the random vector (Y, X) has mean vector 0 and covariance matrix $Cov(Y, X) = \Sigma_{yx} = {\Sigma_{xy}}^T$ and $Cov(X) = \Sigma_{xx}$ is nonsingular, then for any positive-definite matrix Γ, the mxr matrix A and rxn matrix B, for $r \leq min(m, n)$ that minimize

$$tr(E(\Gamma^{1/2}(Y - ABX)(Y - ABX)^T\Gamma^{1/2}))$$

are given by

$$A = \Gamma^{1/2}V, \quad B = V^T\Gamma^{1/2}\Sigma_{yx}{\Sigma_{xx}}^{-1},$$

where $V = (V_1, ..., V_r)$ and V_j is the (normalized) eigenvector that corresponds to the jth largest eigenvalue λ_j^2 of the matrix $\Gamma^{1/2}\Sigma_{yx}{\Sigma_{xx}}^{-1}\Sigma_{xy}\Gamma^{1/2}$, $j = 1, ..., r$.

Eckart-Young Theorem Let S be a matrix of mxn and of rank m, the the Euclidean form $tr((S - P)(S - P)^T)$ is minimum among matrices P of the same size but of rank $r(\leq m)$, when $P = MM^TS$, where M is mxr and the columns of M are the the first r (normalized) eigenvectors of SS^T (i.e. the normalized eigvenvectors).

Remark - SVD (Singular Value Decomposition) Brillinger's theorem is proved by setting $S^* = \Gamma^{1/2}\Sigma_{yx}{\Sigma_{xx}}^{-1}$ (and $P^* = \Gamma^{1/2}AB{\Sigma_{xx}}^{1/2}$), where the positive square roots of SS^T are called the *singular values* of the matrix S.

In general, a mxn matrix S, of rank s, can be expressed in the *singular value decomposition* as

$$S = V\Lambda U^T,$$

where $\Lambda = diag(\lambda_1, ...\lambda_s)$ with ${\lambda_1}^2 \geq ... \geq {\lambda_s}^2 > 0$ being the nonzero eigenvalues of SS^T, $V = (V_1, ...V_s)$ is a mxs matrix s.t. $V^T V = I_s$, and $U = (U_1, ...U_s)$ is a nxs matrix s.t. $U^T U = I_s$.

Back to the model, recall that in the full-rank case, the OLS estimator is given as

$$\hat{\beta}_{OLS} = C^T = {(X^T X)}^{-1} X^T Y,$$

in the reduced-rank case and at the simple level, the estimator can be written as

$$\hat{\beta}_{RR} = B^T A^T = {(X^T X)}^{-1} X^T Y V V^T = \hat{\beta}_{OLS} V V^T.$$

Note.

1. $C = AB = \Gamma^{1/2} V V^T \Gamma^{1/2} Y^T X {(X^T X)}^{-1} = P_\Gamma Y^T X {(X^T X)}^{-1}$, where P_Γ is an idempotent matrix for any Γ (Brillinger's theorem).
2. In the reduced-rank regression, Γ is typically set to be the identity matrix I.

Next let $Y = (Y_1, ..., Y_T)$ and $X = (X_1, ...X_T)$, where T denotes the number of vector observations, and assume that ε_k are iid $\sim N(0, \Sigma_{\varepsilon\varepsilon})$, then the likelihood can be obtained, as shown in the following section.

Maximizing the log-likelihood

$$l(C, \Sigma_{\varepsilon\varepsilon}) = (\frac{T}{2})(log|{\Sigma_{\varepsilon\varepsilon}}^{-1}| - tr({\Sigma_{\varepsilon\varepsilon}}^{-1} W))$$

is the same as minimizing $|W|$ (the determinant of W) and hence minimizing $|\tilde{\Sigma}^{-1}_{\varepsilon\varepsilon} W|$, where $W = (1/T)(Y - CX)(Y - CX)^T$.

W can be further rewritten as

$$\begin{aligned} W &= \frac{1}{T}(Y - \tilde{C}X + (\tilde{C} - AB)X){(Y - \tilde{C}X + (\tilde{C} - AB)X)}^T \\ &= \frac{1}{T}(Y - \tilde{C}X){(Y - \tilde{C}X)}^T + \frac{1}{T}(\tilde{C} - AB) X X^T {(\tilde{C} - AB)}^T \\ &= \tilde{\Sigma_{\varepsilon\varepsilon}} + (\tilde{C} - AB)\hat{\Sigma}_{xx}{(\tilde{C} - AB)}^T. \end{aligned}$$

It has been shown that the solutions

$$\hat{A}_r = \Gamma^{1/2} V, \quad \hat{B}_r = V^T \Gamma^{1/2} \Sigma_{yx} {\Sigma_{xx}}^{-1},$$

with the choice of $\Gamma = \tilde{\Sigma}^{-1}_{\varepsilon\varepsilon}$, minimizes simultaneously all the eigenvalues of $\tilde{\Sigma}^{-1}_{\varepsilon\varepsilon} W$ and hence minimizes $|W|$ (Robinson, 1974). This is because $|\tilde{\Sigma}^{-1}_{\varepsilon\varepsilon} W|$ can be rewritten as

$$|\tilde{\Sigma}^{-1}_{\varepsilon\varepsilon} W| = |I_m + \tilde{\Sigma}^{-1}_{\varepsilon\varepsilon}(\tilde{C} - AB)\hat{\Sigma}_{xx}{(\tilde{C} - AB)}^T| = \prod(1 + \delta_i^2),$$

where δ_i^2 are the eigenvalues of the matrix $\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}(\tilde{C}-AB)\hat{\Sigma}_{xx}(\tilde{C}-AB)^T$. As a result, minimization of $|\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}W|$ is the same as simultaneously minimizing all the eigenvalues of

$$\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}(\tilde{C}-AB)\hat{\Sigma}_{xx}(\tilde{C}-AB)^T \equiv (S-P)(S-P)^T,$$

with $S^* = \tilde{\Sigma}_{\varepsilon\varepsilon}^{-1/2}\tilde{C}\hat{\Sigma}_{xx}^{1/2}$ and $P^* = \tilde{\Sigma}_{\varepsilon\varepsilon}^{-1/2}AB\hat{\Sigma}_{xx}^{1/2}$.

The simultaneous minimization is achieved with P chosen as the rank r approximation of S obtained through the singular value decomposition of S (Lemma 2.2).

Lemma 2.2 Let S be an mxn matrix of rank m and P be an mxn matrix of rank $\leq r(\leq m)$, then, for any i,

$$\lambda_i(S-P) \geq \lambda_{r+i}(S),$$

where $\lambda_i(S)$ denotes the ith largest singular value of the matrix S and $\lambda_{r+i}(S)$ is defined to be zero for $r+i > m$. The equality is attained for all i iff $P = V_r\Lambda_r U_r^T$, where $S = V\Lambda U^T$ represents the singular decomposition of S.

Therefore, $\hat{A}_r$ and $\hat{B}_r$ are the ML estimates for A and B. Or equivalently,

$$\hat{C}_r = \hat{A}_r\hat{B}_r = \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2}V_rV_r^T\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1/2}\tilde{C}$$

is the ML estimate for C.

However, note that

$$\hat{\Sigma_{\varepsilon\varepsilon}} = (1/T)(Y-\hat{C}_rX)(Y-\hat{C}_rX)^T$$

is the reduced-rank ML estimate for $\Sigma_{\varepsilon\varepsilon}$, whereas

$$\tilde{\Sigma_{\varepsilon\varepsilon}} = (1/T)(Y-\tilde{C}X)(Y-\tilde{C}X)^T$$

is the ML estimate in the full-rank case.

Note.

1. $\tilde{C} = \hat{\Sigma}_{yx}\hat{\Sigma}_{xx}^{-1}$ (full-rank) and $\hat{C}_r = \hat{A}_r\hat{B}_r$ (reduced rank).

Discussion

(1) An explicit solution for the matrices A and B can be obtained by computing eigenvalues and eigenvectors of SS^T or S^TS. However, when an explicit solution is not possible to find, iterative procedures need to be considered.

(2) Reduced-rank regression (RRR) model has connections to PCA (principal component analysis) and CCA (canonical correlation analysis). Indeed, the PCA problem can be represented as a RRR model such that

$$Y_k = ABY_k + \varepsilon_k$$

(with $X_k \equiv Y_k$). Setting $\Gamma = I_m$, the solution is given by $A_r = (V_1, ...V_r)$ and $B_r = V^T$ since $\Sigma_{yx} = \Sigma_{yy}$ (Brillinger's theorem).

(3) If the error covariance matrix $\Sigma_{\varepsilon\varepsilon}$ is unknown (or is assumed to have the specified form $\Sigma_{\varepsilon\varepsilon} = \sigma^2\Psi_0$) but is positive-definite, the ML estimates of A and B are obtained by minimizing

$$\begin{aligned}&\frac{1}{\sigma^2}tr({\Psi_0}^{-1}(Y - ABX){(Y - ABX)}^T)\\&= \frac{1}{\sigma^2}tr({\Psi_0}^{-1}(Y - \tilde{C}X){(Y - \tilde{C}X)}^T)+\\&\frac{T}{\sigma^2}tr({\Psi_0}^{-1}(\tilde{C} - AB)\hat{\Sigma}_{xx}{(\tilde{C} - AB)}^T).\end{aligned}$$

Then the ML solution can be found using the Brillinger's theorem.

(4) Reduced-rank regression model focus on the rank reduction of the mean structure CX_k. There are other types of (spatial or spatio-temporal) models such as fixed rank kriging that aim to reduce the rank of the covariance matrix Σ instead.

Reference

Velu, R., & Reinsel, G. C. (2013). Multivariate reduced-rank regression: theory and applications (Vol. 136). Springer Science & Business Media.

Turgeon, M. (2021, January 19). Reduced Rank Regression [Video]. YouTube. https://www.youtube.com/watch?v=AYKmzhbsTwg&list=PLYpxJrn6DQgj3bCPRQDogPKkgzILomlAN&index=4&t=4s.