Abstract

Reduced Rank Regression

The reduced rank regression model is a multivariate regression model with a coefficient matrix with reduced rank. The reduced rank regression algorithm is an estimation procedure, which estimates the reduced rank regression model. It is related to canonical correlations and involves calculating eigenvalues and eigenvectors. We give a number of different applications to regression and time series analysis, and show how the reduced rank regression estimator can be derived as a Gaussian maximum likelihood estimator. We briefly mention asymptotic results.

Reduced Rank Regression

Reduced Rank Regression is an explicit estimation method in multivariate regression, that takes into account the reduced rank restriction on the coefficient matrix.

Reduced rank regression model: We consider the multivariate regression of Y on X and Z of dimension $p,\ q$, and k, respectively: $Y_t = \Pi X_t + \Gamma Z_t + \varepsilon_t,\ t = 1, \ldots, T$. The hypothesis that Π has reduced rank less than or equal to r is expressed as $\Pi = \alpha\beta'$, where α is $p \times r$, and β is $q \times r$, where $r < \min(p, q)$, and gives the reduced rank model

$$Y_t = \alpha\beta' X_t + \Gamma Z_t + \varepsilon_t,\ t = 1, \ldots, T. \tag{1}$$

Reduced rank regression algorithm: In order to describe the algorithm, which we call $RRR(Y, X|Z)$, we introduce the notation for product moments $S_{yx} = T^{-1}\sum_{t=1}^{T} Y_t X_t'$, and $S_{yx.z} = S_{yx} - S_{yz}S_{zz}^{-1}S_{zx}$, etc. The algorithm consists of the following steps:

1. First regress Y and X on Z and form residuals $(Y|Z)_t = Y_t - S_{yz}S_{zz}^{-1}Z_t$, $(X|Z)_t = X_t - S_{xz}S_{zz}^{-1}Z_t$ and product moments

$$S_{yx.z} = T^{-1}\sum_{t=1}^{T}(Y|Z)_t(X|Z)_t' = S_{yx} - S_{yz}S_{zz}^{-1}S_{zx},$$

etc.

2. Next solve the eigenvalue problem

$$|\lambda S_{xx.z} - S_{xy.z}S_{yy.z}^{-1}S_{yx.z}| = 0, \tag{2}$$

where $|.|$ denotes determinant. The ordered eigenvalues are $\Lambda = diag(\lambda_1, \ldots, \lambda_q)$ and the eigenvectors are $V = (v_1, \ldots, v_q)$, so that $S_{xx.z}V\Lambda = S_{xy.z}S_{yy.z}^{-1}S_{yx.z}V$, and V is normalized so that $V'S_{xx.z}V = I_p$ and $V'S_{yx.z}S_{xx.z}^{-1}S_{xy.z}V = \Lambda$. The singular value decomposition provides an efficient way of implementing this procedure, see Doornik and O'Brien (2002)

3. Finally define the estimators

$$= (v_1, \dots, v_r)$$

together with $\hat{\alpha} = S_{yx.z}\hat{\beta}$, and $\hat{\Omega} = S_{yy.z} - S_{yx.z}\hat{\beta}(\hat{\beta}'S_{xx.z}\hat{\beta})^{-1}\hat{\beta}'S_{xy.z}$. Equivalently, once $\hat{\beta}$ has been determined, $\hat{\alpha}$ and $\hat{\Gamma}$ are determined by regression.

The technique of reduced rank regression was introduced by Anderson and Rubin (1949) in connection with the analysis of limited information maximum likelihood and generalized to the reduced rank regression model (1) by Anderson (1951). An excellent source of information is the monograph by Reinsel and Velu (1998), which contains a comprehensive survey of the theory and history of reduced rank regression and its many applications.

Note the difference between the unrestricted estimate $\hat{\Pi}_{OLS} = S_{yx.z}S_{xx.z}^{-1}$ and the reduced rank regression estimate $\hat{\Pi}_{RRR} = S_{yx.z}\hat{\beta}(\hat{\beta}'S_{xx.z}\hat{\beta})^{-1}\hat{\beta}'$ of the coefficient matrix to X.

Various Applications of the Reduced Rank Model and Algorithm

The reduced rank model (1) has many interpretations depending on the context. It is obviously a way of achieving fewer parameters in the possibly large $p \times q$ coefficient matrix Π. Another interpretation is that although X is needed to explain the variation of Y, in practice only a few, r, factors are needed as given by the linear combinations $\beta'X$ in (1).

Restrictions on Π : Anderson (1951) formulated the problem of estimating Π under $p - r$ unknown restrictions $\ell'\Pi = 0$. In (1) these are given by the matrix $\ell = \alpha_\perp$, that is, a $p \times (p - r)$ matrix of full rank for which $\alpha_\perp'\alpha = 0$. The matrix $\alpha_\perp$ is estimated by solving the dual eigenvalue problem $|\lambda S_{yy.z} - S_{yx.z}S_{xx.z}^{-1}S_{xy.z}| = 0$, which has eigenvalues Λ and eigenvectors U, and the estimate is $\hat{\alpha}_\perp = (w_{r+1}, \dots, w_p)$. If $p = q$, we can choose $W = S_{yy.z}^{-1}S_{yx.z}V\Lambda^{-1/2}$.

Canonical correlations: Reduced rank regression is related to canonical correlations, Hotelling (1936). This is most easily expressed if $p = q$, where we find

$$\begin{pmatrix} W & 0 \\ 0 & V \end{pmatrix}' \begin{pmatrix} S_{yy.z} & S_{yx.z} \\ S_{xy.z} & S_{xx.z} \end{pmatrix} \begin{pmatrix} W & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} I_p & \Lambda^{1/2} \\ \Lambda^{1/2} & I_q \end{pmatrix}.$$

This shows that the variables $W'Y$ and $V'X$ are the empirical canonical variates.

Instrumental variable estimation: Let the variables U, V, and X be of dimension p, q, and k respectively with $k \geq q$. Assume that they are jointly Gaussian with mean zero and variance Σ, and that $E(U - \gamma' V)X' = 0$, so that X is an instrument for estimating γ. This means that $\Sigma_{ux} = \gamma'\Sigma_{vx}$, so that

$$E\left(\left(\begin{array}{c} U \\ V \end{array}\right)|X\right) = \left(\begin{array}{c} \gamma'\Sigma_{vx} \\ \Sigma_{vx} \end{array}\right) = \left(\begin{array}{c} \gamma' \\ I_q \end{array}\right)\Sigma_{vx} = \alpha\beta'.$$

It follows that the $(p+q)\times k$ coefficient matrix in a regression of $Y = (U', V')'$ on X has rank q. Thus a reduced rank regression of Y on X is an algorithm for estimating the parameter of interest β using the instruments X. This is the idea in Anderson and Rubin (1949) for the limited information maximum likelihood estimation.

Nonstationary time series: The model

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Gamma\Delta Y_{t-2} + \varepsilon_t, \ t = 1, \ldots, T \tag{3}$$

determines a multivariate times series Y_t, and the reduced rank of $\alpha\beta'$ implies nonstationarity of the time series. Under suitable conditions, see Johansen (1996) Y_t is nonstationary and ΔY_t and $\beta' Y_t$ are stationary. Thus Y_t is a cointegrated time series, see Engle and Granger (1987).

Common features: Engle and Kozicki (1993) used model (3) and assumed reduced rank of the matrix $(\alpha, \Gamma) = \xi\eta'$, so that $\Delta Y_t = \xi\eta'(Y_{t-1}'\beta, \Delta Y_{t-2}')' + \varepsilon_t$. In this case $\xi_\perp'\Delta Y_t = \xi_\perp'\varepsilon_t$ determines a random walk, where the common cyclic features have been eliminated.

Prediction: Box and Tiao (1977) analysed the model $Y_t = \Pi Y_{t-1} + \Gamma Y_{t-2} + \varepsilon_t$, and asked which linear combinations of the current values, $v'Y_t$, are best predicted by a linear combination of the past (Y_{t-1}, Y_{t-2}), and hence introduced the analysis of canonical variates in the context of prediction of times series.

The Gaussian Likelihood Analysis

If the errors ε_t in model (1) are i.i.d. Gaussian $N_p(0, \Omega)$, and independent of $\{X_s, Z_s, s \leq t\}$, the (conditional or partial) Gaussian likelihood is

$$-\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - \alpha\beta' X_t - \Gamma Z_t)'\Omega^{-1}(Y_t - \alpha\beta' X_t - \Gamma Z_t)$$

Anderson (1951) introduced the RRR algorithm as a calculation of the maximum likelihood estimator of $\alpha\beta'$. The Frisch-Waugh Theorem shows that one can partial out the parameter Γ by regression, as in the first step of the algorithm. We next regress $(Y|Z)$ on $(\beta' X|Z)$ and find estimates of α and Ω, and the maximized likelihood function as functions of β :

$$\begin{aligned}\hat{\alpha}(\beta) &= S_{yx.z}\beta(\beta' S_{xx.z}\beta)^{-1}, \\ \hat{\Omega}(\beta) &= S_{yy.z} - S_{yx.z}\beta(\beta' S_{xx.z}\beta)^{-1}\beta' S_{xy.z}, \\ L_{\max}^{-2/T}(\beta) &= |\hat{\Omega}(\beta)|.\end{aligned} \tag{4}$$

The identity

$$\begin{aligned}&\left|\begin{pmatrix} S_{yy.z} & S_{yx.z}\beta \\ \beta' S_{xy.z} & \beta' S_{xx.z}\beta \end{pmatrix}\right| = |S_{yy.z}||\beta' S_{xx.z}\beta - \beta' S_{xy.z}S_{yy.z}^{-1}S_{yx.z}\beta| \\ &= |\beta' S_{xx.z}\beta||S_{yy.z} - S_{yx.z}\beta(\beta' S_{xx.z}\beta)^{-1}\beta' S_{xy.z}| = |\beta' S_{xx.z}\beta||\hat{\Omega}(\beta)|,\end{aligned}$$

shows that $L_{\max}^{-2/T}(\beta) = |S_{yy.z}||\beta'(S_{xx.z} - S_{xy.z}S_{yy.z}^{-1}S_{yx.z})\beta|/|\beta' S_{xx.z}\beta|$ so that β has to be chosen to minimize this. Differentiating with respect to β we find that β has to satisfy the relation $S_{xx.z}\beta = S_{xy.z}S_{yy.z}^{-1}S_{yx.z}\beta\xi$ for some $r \times r$ matrix ξ. This shows, see Johansen (1996), that the space spanned by the columns of β, is spanned by r of the eigenvectors of (2), and hence, that choosing the largest λ_i gives the smallest value of $L_{\max}^{-1/T}(\beta)$, so that

$$\hat{\beta} = (v_1, \ldots, v_r), \quad L_{\max}^{-2/T}(\hat{\beta}) = |S_{yy.z}|\prod_{i=1}^{r}(1 - \lambda_i).$$

Hypothesis testing: The test statistic for rank of Π can be calculated from the eigenvalues because the eigenvalue problem solves the maximization of the likelihood for all values of r simultaneously. The likelihood ratio test statistic for the hypothesis $\text{rank}(\Pi) \leq r$, as derived by Anderson (1951), is

$$-2\log LR(\text{rank}(\Pi) \leq r) = T\sum_{i=r+1}^{\min(p,q)} \log(1 - \lambda_i). \tag{5}$$

Bartlett (1938) suggested using this statistic to test that r canonical correlations between Y and X were zero and hence that Π had reduced rank.

The simplest hypothesis to test on β is $\beta = H\phi$. We can estimate β under this restriction by $RRR(Y, H'X|Z)$ and therefore calculate the likelihood ratio statistic using reduced rank regression. If, on the other hand, we have restrictions on the individual vectors $\beta = (H_1\phi_1, \dots, H_r\phi_r)$, then reduced rank regression does not provide a maximum likelihood estimator, but we can switch between reduced rank regressions as follows: For fixed $\beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_r$, we can find an estimator for ϕ_i and hence $\beta_i = H_i\phi_i$ by

$$RRR(Y, H_i'X|Z, (\beta_1, \dots \beta_{i-1}, \beta_{i+1}, \dots, \beta_r)'X).$$

By switching between the vectors in β, we have an algorithm which is useful in practice and which maximizes the likelihood in each step.

A switching algorithm: Another algorithm, that is useful for this model, is to consider the first order condition for β, when Γ has been eliminated, which has solution

$$\hat{\beta}(\alpha, \Omega) = S_{xx.z}^{-1} S_{xy.z} \Omega^{-1} \alpha \left(\alpha' \Omega^{-1} \alpha\right)^{-1}.$$

Combining this with (4) suggests a switching algorithm:

First choose some initial estimator $\hat{\beta}_0$, then switch between estimating α and Ω for fixed β by least squares, and estimating β for fixed α and Ω by generalized least squares.

This switching algorithm maximizes the likelihood function in each step and any limit point will be a stationary point. It seems to work well in practice. There are natural hypotheses one can test in the reduced rank model, like general linear restrictions on β, which are not solved by the reduced rank regression algorithm, whereas the above algorithm can be modified to give a solution.

Asymptotic distributions in the stationary case: The asymptotic distributions of the estimators and test statistics can be described under the assumption that the process (Y_t, X_t, Z_t) is stationary with finite second moments. It can be shown that estimators are asymptotically Gaussian and test statistics for hypotheses both for rank and for β are asymptotically χ^2, see Robinson (1973).

Asymptotic distributions in the nonstationary case: If the processes are nonstationary a different type of asymptotics is needed. As an example consider model (3) for $I(1)$ variables. When discussing the asymptotic distribution of the estimators, the normalization by $\hat{\beta}' S_{xx.z} \hat{\beta} = I_r$ is not convenient, and it is necessary to identify the vectors differently.

One can then prove, see Johansen (1996), that the estimates of α, Γ, and Ω are asymptotically Gaussian and have the same limit distribution as if β

were known, that is, the asymptotic distribution they have in the regression of ΔY_t on the stationary variables $\beta' Y_{t-1}$ and ΔY_{t-1}.

The asymptotic distribution of $\hat{\beta}$ is mixed Gaussian, where the mixing parameter is the (random) limit of the observed information. Therefore, by normalizing on the observed information, we obtain asymptotic χ^2 inference for hypotheses on β.

The limit distribution of the likelihood ratio test statistic for rank, see (5), is given by a generalization of the Dickey-Fuller distribution:

$$DF_{p-r} = tr\{\int_0^1 (dW)W' \left(\int_0^1 WW'du\right)^{-1} \int_0^1 W(dW)'\},$$

where W is a standard Brownian motion in $p-r$ dimensions. The quantiles of this distribution can at present only be calculated by simulation if $p-r > 1$. The limit distribution has to be modified if deterministic terms are included in the model.

Søren Johansen,

JEL Keywords: C10, C13

Bibliography

Anderson, T.W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327-351.

Anderson, T.W. and Rubin, H. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46-63.

Bartlett, M.S. 1938. Further aspects of the theory of multiple regression. Proceedings of the Cambridge Philosophical Society 34, 33-40.

Box, G.E.P. and Tiao, G.C. 1977. A canonical analysis of multiple time series. Biometrika 64, 355-365.

Doornik, J.A. and O'Brien, R.J. 2002. Numerically stable cointegration analysis. Computational Statistics and Data Analysis 41, 185-193.

Engle, R.F. and Granger, C.W.J. 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251-276.

Engle, R.F. and Kozicki, S. 1993. Testing for common factors (with comments). Journal of Business Economics & Statistics 11,369-278.

Hotelling, H. 1936. Relations between two sets of variables. Biometrika 28, 321-377.

Johansen, S. 1996. Likelihood Based Inference on Cointegration in the Vector Autoregressive Model. Oxford: Oxford University Press.

Reinsel, G.C. and Velu, R.P. 1998. Multivariate Reduced rank regression, Lecture Notes in Statistics. New York: Springer.

Robinson. P.M. 1973. Generalized canonical analysis for time series. Journal of Multivariate Analysis 3, 141-160.