Joint Redundancy Analysis by a multivariate linear predictor

Analisi di ridondanza condivisa sulla base di un predittore lineare multivariato

Laura Marcis, Renato Salvatore

Abstract A common multi-group Redundancy Analysis is introduced, when the reduced space is given by a singular value decomposition of a multivariate best linear predictor. The algorithm finds a nearby *OLS* fixed-effects estimates by a least squares closed-form solution, provided by the standardized predictor. The empirical predictor is given by an extension of the distribution-free variance least squares method to an iterative multivariate response algorithm.

Abstract *Il lavoro introduce una Analisi di Ridondanza sulla base di gruppi indipendenti, utilizzando la decomposizione ai valori singolari di un predittore lineare multivariato. L'algoritmo fornisce stime di effetti fissi vicini alle stime OLS, attraverso una soluzione esatta sulla base del predittore standardizzato. La stima del predittore empirico è basata sull'estensione del metodo ai minimi quadrati della varianza del modello al caso multivariato, seguendo un approccio iterativo.*

Key words: Redundancy analysis, linear mixed model, empirical best linear unbiased predictor, variance least squares

Laura Marcis
Department of Economics and Law, University of Cassino and Southern Latium, e-mail: laura.marcis@unicas.it

Renato Salvatore
Department of Economics and Law, University of Cassino and Southern Latium e-mail: rsalvatore@unicas.it

1 Introduction

Redundancy Analysis (*RDA*) was originally introduced [7] in order to capture the effects on a reduced space of the linear dependence by a set of criterion variables **Y** from a set of predictors **X**. Partial *RDA*, constrained *RDA* and ridge-type regularized *RDA* were also introduced, where the goal is substantially of two types: firstly, to highlight the effects of a subset of some conditioning predictors [2], to remove, and, secondly, to assess a ridge estimator to reduce the mean squared error of the multivariate regression by some nearby collinear predictors [6]. Even though the *RDA* provides a constrained analysis of the whole linear relations between the two sets of variables, and an unconstrained analysis given by the set of multivariate regression residuals, it is straightforward to relate *RDA* with principal component analysis (*PCA*), see for example [4]. One of research issues in the field of the *PCA* is the simultaneous *PC*-reduction of a set of independent groups of observations, collected by a multivariate random vector **Y** ([5],[1]). A general linear mixed model [3] is usually employed to represent the relationship between the sets of criterion and predictor variables, when the goal is to predict a specific group (subject) contribution to the linear dependence. For this reason, a *RDA* of the predicted criterion variables by the best linear unbiased predictor at group level may be quite representative into this contribution. Further, it is also useful to perform *RDA* of the modeled predicted data to investigate the "common" groups redundancy on the criterion variables, and on the multivariate mixed model conditioned residuals. This paper introduces a joint *RDA* by a least-squares solution for an optimal fixed-effects estimate from the data collected by the linear mixed model predictors of the dependent variables. The singular value decomposition of the resulting linear regression model estimates gives the best projection in the common reduced subspace of the best unbiased predictor by the whole set of random effects. The application uses an extension to the multivariate case of the variance least squares algorithm to estimate a variance components MANOVA model for data repeated over time.

2 Joint Redundancy Analysis. Estimation of model parameters

Given a q-variate random vector **Y**, consider the case when **Y** is partitioned in n subjects (groups), each of them with n_i individuals. We assume that the population model for the n subjects is $\mathbf{y}_{i|q\times 1} = \mathbf{B}'_{q\times p}\mathbf{x}_{i|p\times 1} + \mathbf{a}_{i|q\times 1}$, where **B** is the matrix of fixed regression coefficients, and $\mathbf{a}_i \sim N(0, \Sigma_a)$ is a q-variate random effect. Given a sample of N units (repeated measurements), then the multivariate random effects model assumes the general structure $\mathbf{Y}_{N\times q} = \mathbf{X}^{+}_{N\times p}\mathbf{B}_{p\times q} + \mathbf{Z}^{+}_{N\times n}\mathbf{A}_{n\times q} + \mathbf{E}_{N\times q}$, with $\mathbf{X}^{+}$ the matrix of data covariates, $\mathbf{Z}^{+}$ the design matrix of random effects, and $\mathbf{E}$ the matrix of regression within-subject errors. Further both $\mathbf{Y}$ and $\mathbf{X}^{+}$ are assumed as columnwise centered and standardized. Rewriting the last model in the vector form, with $\mathbf{y}^* = vec(\mathbf{Y})$, $\mathbf{X} = (\mathbf{I}\otimes\mathbf{X}^{+})$, $\beta = vec(\mathbf{B})$, $\mathbf{Za} = vec(\mathbf{ZA}) = (\mathbf{I}\otimes\mathbf{Z})vec(\mathbf{A})$, and given for the sake of simplicity a balanced design ($n_i = k$), the multivariate linear best predictor is given by $\widetilde{\mathbf{y}}^* = vec(\widetilde{\mathbf{Y}}) = \mathbf{X}\widehat{\beta} + \mathbf{Z}\widetilde{\mathbf{a}} = \left\{\mathbf{I} - (\Sigma_a \otimes \mathbf{ZZ}')cov(\mathbf{y}^*)^{-1}\right\}\mathbf{X}\widehat{\beta} + (\Sigma_a \otimes \mathbf{ZZ}')cov(\mathbf{y}^*)^{-1}\mathbf{y}^* = \Gamma\mathbf{y}^* + (\mathbf{I}-\Gamma)\mathbf{X}\widehat{\beta}$.

Here $\widehat{\beta} = \widehat{\beta}_{gls}$, $\mathbf{Z}\widetilde{\mathbf{a}} = (\Sigma_a \otimes \mathbf{ZZ}')cov(\mathbf{y}^*)^{-1}(\mathbf{y}^* - \mathbf{X}\widehat{\beta})$, $cov(\mathbf{y}^*) = (\mathbf{I}\otimes\mathbf{Z})\ (\Sigma_a \otimes \mathbf{I}_k)(\mathbf{I}\otimes\mathbf{Z}') + cov(vec(\mathbf{E})) = (\Sigma_a \otimes \mathbf{ZZ}') + (\Sigma_e \otimes \mathbf{I}_n)\otimes\mathbf{I}_k$, $\Gamma = (\Sigma_a \otimes \mathbf{ZZ}')cov(\mathbf{y}^*)^{-1}$.

By standard *Redundancy Analysis* (*RDA*), a reduced-rank subspace is given by a *singular value decomposition* (*SVD*) of the multivariate regression predicted values $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{B}}_{ols}$. To get a reduced subspace by the multivariate linear best predictor $\widetilde{\mathbf{Y}}$, thus we set $\widetilde{\mathbf{Y}} = \mathbf{X}^{+}\widehat{\mathbf{B}} + \mathbf{Z}\widetilde{\mathbf{A}} = \mathbf{U}_{\widetilde{Y}}\Lambda_{\widetilde{Y}}\mathbf{V}'_{\widetilde{Y}}$ as a possible *SVD* representation of the redundant information in the criterion variables, captured by the dispersion matrix $\Sigma_{\mathbf{y}^*|\mathbf{X}} = var(\mathbf{y}^*) - cov(\mathbf{y}^*,\mathbf{X})var(\mathbf{X})^{-1}cov(\mathbf{X},\mathbf{y}^*)$. Even though a joint-subject reduction subspace is given by the fixed-effects model estimates $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{B}}_{gls}$, we are interested in the simultaneous representation of all the predicted $\widetilde{\mathbf{a}}_i$'s, given by a common projection subspace. To do this, we find the matrix of fixed effects $\mathbf{B}$ closest to the fixed $\widehat{\mathbf{B}}_{ols}$, by setting the minimum Froebenius norm by the multivariate predictor $\widetilde{\mathbf{Y}}$, of the difference $\digamma = \widetilde{\mathbf{Y}}^{**}var(\widetilde{\mathbf{y}})^{-\frac{1}{2}} - \mathbf{X}^{+}(\mathbf{B}), \widetilde{\mathbf{Y}}^{**} = \widetilde{\mathbf{Y}} - E(\widetilde{\mathbf{Y}}) = \widetilde{\mathbf{Y}} - \mathbf{1}_N E[(\widetilde{\mathbf{y}}_{|q\times 1})']$: $\|\digamma\|^2 = tr(\mathbf{f}'\mathbf{f}) = \left\|\widetilde{\mathbf{Y}}^{**}\Sigma^{-\frac{1}{2}} - \mathbf{X}^{+}(\mathbf{B})\right\|^2 = \min$.

Note that $var(\widetilde{\mathbf{y}}) = \Sigma = E\left\{(\widetilde{\mathbf{Y}}-\mathbf{Y})'(\widetilde{\mathbf{Y}}-\mathbf{Y})\right\}$, with $(\widetilde{\mathbf{Y}}-\mathbf{Y})'(\widetilde{\mathbf{Y}}-\mathbf{Y})$ the random matrix of $q\times q$ cross products $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}$ given on the basis of the subjects covariances $cov(\widetilde{\mathbf{y}}_i^* - \mathbf{y}_i^*) = \mathbf{X}_i(cov\widehat{\beta}_{gls})\mathbf{X}_i' + \mathbf{Z}_i cov(\widetilde{\mathbf{a}}_i - \mathbf{a}_i)\mathbf{Z}_i' + cov(vec(\mathbf{E}))$.

Now, setting $\mathbf{f} = vec(\digamma) = (\Sigma^{-\frac{1}{2}}\otimes\mathbf{I}_N)\widetilde{\mathbf{y}}^{**} - (\mathbf{I}_q\otimes\mathbf{X}^{+})\overline{\beta} = \overline{\Sigma}^{-\frac{1}{2}}\widetilde{\mathbf{y}}^{**} - \mathbf{X}\overline{\beta}$, $\overline{\beta} = vec(\mathbf{B})$, $\overline{\Sigma}^{-\frac{1}{2}} = (\Sigma^{-\frac{1}{2}}\otimes\mathbf{I}_N)$, we come the following properties of $\overline{\beta}$: $tr(\mathbf{f}'\mathbf{f}) = tr\left\{(\overline{\Sigma}^{-\frac{1}{2}}\widetilde{\mathbf{y}}^{**} - \mathbf{X}\overline{\beta})'(\overline{\Sigma}^{-\frac{1}{2}}\widetilde{\mathbf{y}}^{**} - \mathbf{X}\overline{\beta})\right\} = (\widetilde{\mathbf{y}}^{**} - \overline{\mathbf{X}}\overline{\beta})'\overline{\Sigma}^{-1}(\widetilde{\mathbf{y}}^{**} - \overline{\mathbf{X}}\overline{\beta})$, where $\overline{\mathbf{X}} = \overline{\Sigma}^{\frac{1}{2}}\mathbf{X}$. Thus, $\widehat{\overline{\beta}} = (\overline{\mathbf{X}}'\overline{\Sigma}^{-1}\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\overline{\Sigma}^{-1}\widetilde{\mathbf{y}}^{**}$ is the q-variate vector in the subspace spanned by the columns of the matrix $\overline{\mathbf{X}}$, with $\widetilde{\mathbf{y}}^{**}$ orthogonal to the columns of $\overline{\mathbf{X}}$ in the metric of $\overline{\Sigma}^{-1}$, $\widetilde{\mathbf{y}}^{**\prime}\overline{\Sigma}^{-1}\overline{\mathbf{x}} = 0$. Then, $P_{\overline{X}} = \overline{\mathbf{X}}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'$ is the projection matrix of the predictor $\widetilde{\mathbf{y}}^{**}$ onto the joint subspace by $\overline{\mathbf{X}}$. The *SVD* of $\overline{\widetilde{\mathbf{Y}}}^{**} = \mathbf{X}^{+}\widehat{\overline{\beta}}$

gives the common rescaled predictor's coordinates, $\mathbf{U}_{\widetilde{\widetilde{Y}}}\Lambda_{\widetilde{\widetilde{Y}}}\mathbf{V}'_{\widetilde{\widetilde{Y}}}$, further noticing that $\mathbf{U}^*_{\widetilde{Y}} = \widetilde{\mathbf{Y}}\mathbf{V}^{-1}\Lambda_{\widetilde{\widetilde{Y}}}$ contains the row joint reduced coordinates in the space of $\widetilde{\mathbf{Y}}$.

In order to avoid distributional assumptions for the multivariate data vector $\mathbf{Y}$, we introduce an Iterative multivariate Variance Least Squares (*IVLS*) estimation. The objective function to minimize is $VLS = trace(\Xi - \mathbf{U} - \mathbf{D})^2$, with $\Xi_{|N\times N}$ the empirical model covariance matrix. The algorithm is based on alternating least squares in a two-step iterative optimization process. At every iteration IVLS first fixes $\mathbf{U}$ and solves for $\mathbf{D}$, and following that it fixes $\mathbf{D}$ and solves for $\mathbf{U}$. Since *LS* solution is unique, in each step the *VLS* function can either decrease or stay unchanged, but never increase. Alternating between the two steps iteratively guarantees convergence only to a local minima, because it ultimately depends on the initial values for $\mathbf{U}$. The iterations are related to the following steps: a) from the group covariance matrices $\mathbf{U}$, first minimize VLS to obtain the estimates of $\mathbf{D}$, where Ξ is given by the multivariate *OLS* cross-products of residuals; b) after the estimation of the matrix $\widehat{\mathbf{B}}_{GLS}$, minimize VLS, setting the same error covariance matrix among groups, and c), Iterate a) and b), until convergence to the minimum. The number of iterations may vary by the choice of the specific model random effects and error covariance matrices. Applications of the *Joint RDA* may be related to different types of available data, and then accommodate a variety of patterned covariance matrices. Further, groups can be dependent or independent, even in space, time, and space-time correlated data. The *IVLS* estimator at each step is unbiased, by the following Lemma.

Lemma *(Unbiasedness of the IVLS estimator)*. Under the balanced p -variate variance components model $\mathbf{Y} = \mathbf{XB} + \mathbf{ZA} + \mathbf{E}$, with covariance matrix $\mathbf{D} + \mathbf{U}$, $\mathbf{D} = (\mathbf{I} \otimes \mathbf{Z})cov(vec(\mathbf{A}))(\mathbf{I} \otimes \mathbf{Z}')$, $\mathbf{U} = cov(vec(\mathbf{E}))$, and known matrix $\mathbf{U}$, for the *IVLS* estimator of the parameters θ in $\mathbf{D}$ we have $E\left\{\mathbf{D} = D(\widehat{\theta}_{IVLS})\right\} = D(\theta)$.

3 Application and concluding remarks

Recent national laws reformed the Italian Budget law, provided that the "Benessere Equo e Sostenibile (BES) - Fair and Sustainable Well-being (FSW)" [8] indicators should contribute to define those economic policies which largely affect some fundamental dimensions for the quality of life. The Italian Statistical Institute provide these indicators annually. The Ministry of Finance and Economics most recent publication is the Budget Law 2019 where it is possible to find the trend and programmatic forecasts relating to the 12 FSW indicators and the analysis of most recent trends, at the levels NUTS2 and NUTS3. We analyze 5 of the 12 FSW indicators available in the years 2013-2016 (4 time istants), at the level of NUTS2. The random multivariate vector is partitioned in repeated observations of the same administrative Region of Italy in the 4 time instants. We take in consideration a balanced multivariate Mixed *MANOVA* Model (*MMM*), with an *AR(1)* error stucture: $\mathbf{Y}_{|mk\times p} = \mathbf{X}_{|mk\times l}\mathbf{B}_{|l\times p} + \mathbf{Z}_{|mk\times pm}\mathbf{A}_{|pm\times p} + \mathbf{E}_{|mk\times p}$, where $p = 5$, $m = 20$, $k = t = 4$, that is a balanced model, with a random intercept and an *AR(1)* error. Then: $vec(\mathbf{Y}) = (\mathbf{I}_p \otimes 1_{mt})vec(\mathbf{B}) + (\mathbf{I}_p \otimes Z)vec(\mathbf{A}) + vec(\mathbf{E})$; $\mathbf{y}^* = vec(\mathbf{Y}), \mathbf{X}^* = (\mathbf{I}_p \otimes \mathbf{X}) = (\mathbf{I}_p \otimes \mathbf{1}_{mt})$, $\beta^* = vec(\mathbf{B})$, $\mathbf{Z}^*\mathbf{a}^* = (\mathbf{I}_p \otimes \mathbf{Z})vec(\mathbf{A})$.

Further we have $cov(\mathbf{y}^*) = (\mathbf{I}_p \otimes \mathbf{I}_m \otimes \mathbf{1}_k)(\Sigma_a \otimes \mathbf{I}_m)(\mathbf{I}_p \otimes \mathbf{I}_m \otimes \mathbf{1}'_k) + cov(vec(\mathbf{E})) = \Sigma_a \otimes (\mathbf{I}_m \otimes \mathbf{1}_k \mathbf{1}'_k) + (\Sigma_e \otimes \mathbf{I}_n) \otimes \Omega$. Finally, after the iterative *VLS* estimation, the predictor is given by $\widetilde{\mathbf{y}}^* = \mathbf{X}^*\widehat{\beta}^*_{GLS} + \mathbf{Z}^*\widetilde{\mathbf{a}}^* = \Gamma \mathbf{y}^* + (\mathbf{I} - \Gamma)\mathbf{X}^*\widehat{\beta}^*_{GLS}$, $\Gamma = (\Sigma_a \otimes \mathbf{Z}\mathbf{Z}')cov(\mathbf{y}^*)^{-1}$. Note that the matrix Γ specifies both the contribution of the regression model and the observed data to the prediction.

We assume equicorrelation both of the multivariate random effects and the residual covariance, together with the *AR(1)* structure of the error:

$$\Sigma_a = \sigma_a^2 \times \begin{bmatrix} 1 & \rho_a & \cdots & \rho_a \\ \rho_a & 1 & \cdots & \rho_a \\ \vdots & \cdots & \ddots & \vdots \\ \rho_a & \rho_a & \cdots & 1 \end{bmatrix}_{5\times 5} \Sigma_e = \sigma_e^2 \times \begin{bmatrix} 1 & \rho_e & \cdots & \rho_e \\ \rho_e & 1 & \cdots & \rho_e \\ \vdots & \cdots & \ddots & \vdots \\ \rho_e & \rho_e & \cdots & 1 \end{bmatrix}_{5\times 5} \Omega = \frac{1}{1-\rho_t^2} \begin{pmatrix} 1 & \rho_t & \rho_t^2 & \rho_t^3 \\ \rho_t & 1 & \rho_t & \rho_t^2 \\ \rho_t^2 & \rho_t & 1 & \rho_t \\ \rho_t^3 & \rho_t^2 & \rho_t & 1 \end{pmatrix}_{4\times 4}$$

Figure 1 reports a comparison between observed and predicted data. Bold lines refer to predicted loadings and black dots are the predicted scores. Dashed lines and grey dots come from the standard *PCA*.

In conclusion, the paper introduces *RDA* of a multivariate predictor to perform a common survey of the predicted data, a joint *RDA* analysis. Given a multivariate vector with independent groups, and a random effects population model, the joint *RDA* relies on the assumption that the linear model itself is able to predict accurately specific subjects or group representatives, even in time and spatial dependent data. After using a linear mixed model, the joint *RDA* explores data that originates in part from regressive process and in part from the observed, to understand the contribution to the linear dependence of the observed and of predictions. We suggest the use of this approach when the research issues are related to the use of model covariates and

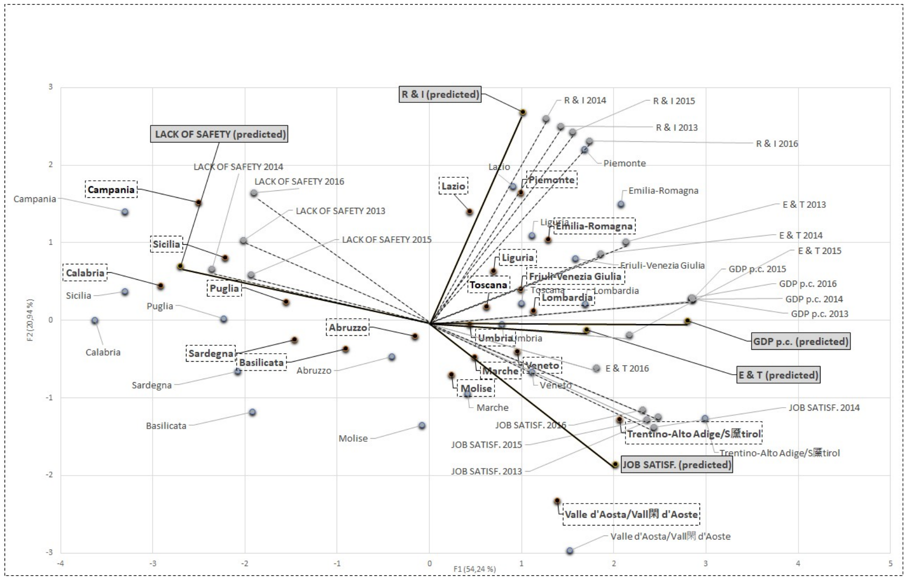

Fig. 1 Multiple Factor Analysis (MFA), observed factor loadings and scores per year (in grey); predicted loadings and scores (in black).

specific patterned covariance matrices. Further, the impact of choosing the model structure is easily recognizable when we investigate changes in the data description by the common factors.

References

1. Bechger, T.M., Blanca, M.J., Maris, G.: The analysis of multivariate group differences using common principal components. Structural Equation Modelling **21**, 577-587 (2014)
2. Borcard, D., Legendre, P., Drapeau, P.: Partialling out the spatial component of ecological variation. Ecology **84**, 511-525 (1992)
3. Demidenko: Mixed Models: theory and applications. Wiley (2004)
4. Härdle, W.K., Simar, L.: Principal Components Analysis. In: Applied Multivariate Statistical Analysis, pp. 319-358. Springer Berlin Heidelberg (2015)
5. Neuenschwander, B.E., Flury, B.D.: Common Principal Components for Dependent Random Vectors. J. of Multivar. Analysis **75**, 163-183 (2000)
6. Takane, Y., Hwang, H.: Regularized linear and kernel redundancy analysis. Computational Statistics and Data Analysis. **52(1)** 394-405 (2007)
7. van den Wollenberg, A.L.: Redundancy analysis an alternative for canonical correlation analysis. Psychometrika **42**, 207–219 (1977)
8. https://www.istat.it/en/