



%ERA: A SAS Macro for Extended Redundancy Analysis

Pietro Giorgio Lovaglio
University of Milano-Bicocca

Gianmarco Vacca
University of Milano-Bicocca

Abstract

A new approach to structural equation modeling based on so-called extended redundancy analysis has been recently proposed in the literature, enhanced with the added characteristic of generalizing redundancy analysis and reduced-rank regression models for more than two blocks. In this approach, the relationships between the observed exogenous variables and the observed endogenous variables are moderated by the presence of unobservable composites that were estimated as linear combinations of exogenous variables, permitting a great flexibility to specify and fit a variety of structural relationships. In this paper, we propose the SAS macro %ERA to specify and fit structural relationships in the extended redundancy analysis (ERA) framework. Two examples (simulation and real data) are provided in order to reproduce results appearing in the original article where ERA was proposed.

Keywords: extended redundancy analysis, SAS macro, alternating least squares, latent components.

1. Introduction

Within the component analysis (CA) framework (Meredith and Millsap 1985; Schonemann and Steiger 1976), redundancy analysis (RA; Van den Wollenberg 1977) is the simplest type of structural-equation model between two sets of observed variables, in which latent variables are intended as components. The aim of RA is to extract a series of linear components from a set of exogenous variables in such a way that they are mutually orthogonal and successively account for the maximum variance of a set of endogenous variables. In this framework, RA may be viewed as a special type of structural-equation model where: (C1) a formative relationship is always assumed between the unobserved and observed (exogenous) variables, and (C2) endogenous variables are always observed ones.

A few attempts have been made to extend RA to more than two sets of variables. However, they are limited to relationships among three sets of variables (Davies and Tso 1982; Reinsel and Velu 1998) as well as being limited to particular types of models (Bougeard, Qannari, Lupo, and Hanafi 2011). Recently, a new method based on so-called extended redundancy analysis (ERA; Takane and Hwang 2005), which generalizes RA for more than two blocks, has been proposed in the literature. In ERA, the relationships between the observed exogenous variables and the observed endogenous variables are moderated by the presence of linear composites (hereinafter LCs): LCs are estimated as exact linear combinations of formative indicators, and both component weights and component loadings are estimated by consistently minimizing a single objective function.

Unfortunately, software to estimate ERA models is not available, hence in the present paper a SAS (SAS Institute Inc. 2011) macro to specify and fit ERA models is proposed. The paper is organized as follows: Section 2 describes the ERA framework, Section 3 presents and discusses the macro %ERA, Section 4 illustrates two studies (empirical and simulative) to assess if the proposed macro is able to reproduce the results of ERA as given in the original paper where the method was proposed, Section 5 offers conclusions.

2. The ERA model: Specification and fitting

The ERA model can generally be stated as follows. Let \mathbf{Y} denote an $n \times p$ matrix consisting of p observed endogenous variables on n subjects. Let \mathbf{X} denote an $n \times q$ matrix consisting of q observed exogenous (formative) variables. Assume that the columns of the matrices are mean centered with variance scaled to unit. Hence, the model can be expressed as

$$\mathbf{Y} = \mathbf{X}\mathbf{W}\mathbf{A}^\top + \mathbf{E} = \mathbf{F}\mathbf{A}^\top + \mathbf{E}, \quad \text{rank}(\mathbf{W}\mathbf{A}^\top) = D \leq \min\{q, p\}, \quad (1)$$

where \mathbf{W} denotes a $q \times D$ matrix of composite weights, \mathbf{A}^\top denotes a $D \times p$ matrix of composite loadings, \mathbf{E} denotes an $n \times p$ matrix of residuals, and \mathbf{F} ($= \mathbf{X}\mathbf{W}$) denotes an $n \times D$ matrix of composite scores. For identification, \mathbf{F} is restricted to be $\text{diag}(\mathbf{F}^\top \mathbf{F}) = \mathbf{I}_D$. To illustrate further, suppose there are for example three sets of variables: $\mathbf{X}_1 = [\mathbf{x}_1 | \mathbf{x}_2]$, $\mathbf{X}_2 = [\mathbf{x}_3 | \mathbf{x}_4]$, and $\mathbf{Y} = [\mathbf{y}_1 | \mathbf{y}_2]$, and suppose a certain relationship among the three sets of variables is considered, as displayed in Figure 1.

The associated ERA model is:

$$\mathbf{Y} = [\mathbf{X}_1 | \mathbf{X}_2] \begin{bmatrix} \mathbf{w}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{w}_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + [\mathbf{e}_1 | \mathbf{e}_2] = \mathbf{X}\mathbf{W}\mathbf{A}^\top + \mathbf{E}, \quad (2)$$

where $\mathbf{w}_1 = (w_1, w_2)$, $\mathbf{w}_2 = (w_3, w_4)$ are composite weights for \mathbf{X}_1 and \mathbf{X}_2 , respectively, $\mathbf{F} = [\mathbf{f}_1 | \mathbf{f}_2] = [\mathbf{x}_1 w_1 + \mathbf{x}_2 w_2 | \mathbf{x}_3 w_3 + \mathbf{x}_4 w_4]$ and $\mathbf{E} = [\mathbf{e}_1 | \mathbf{e}_2]$. Equation 1 reduces to the RA model when no variables are shared between $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$ and \mathbf{Y} , as well as when no constraints other than on the rank of $\mathbf{W}\mathbf{A}^\top$ are imposed on \mathbf{W} and \mathbf{A}^\top (e.g., constrained values such as zeros or ones).

Despite the conditions (C1) and (C2) (given in Section 1 of the CA framework to which ERA belongs), structural-equation models specified in the ERA framework are quite comprehensive, enabling us to specify various and complex situations. For example, in Equation 1, let $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$ be a matrix composed of two blocks, that generate an $n \times 2$ ($= D$) matrix of composite

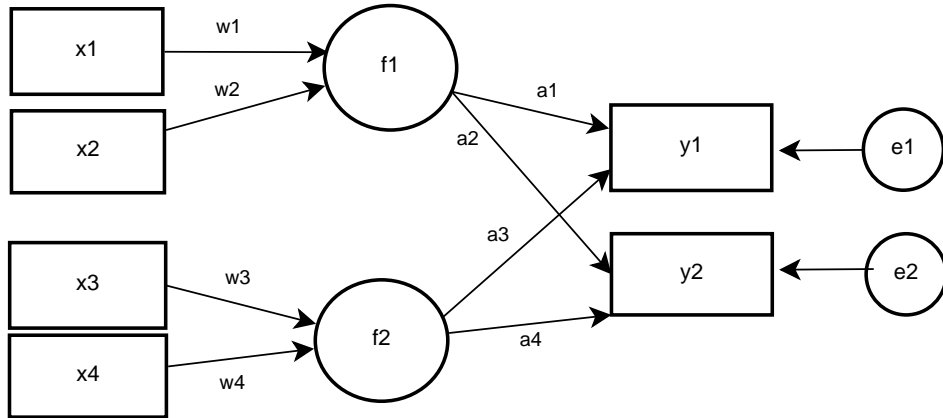


Figure 1: A basic ERA model.

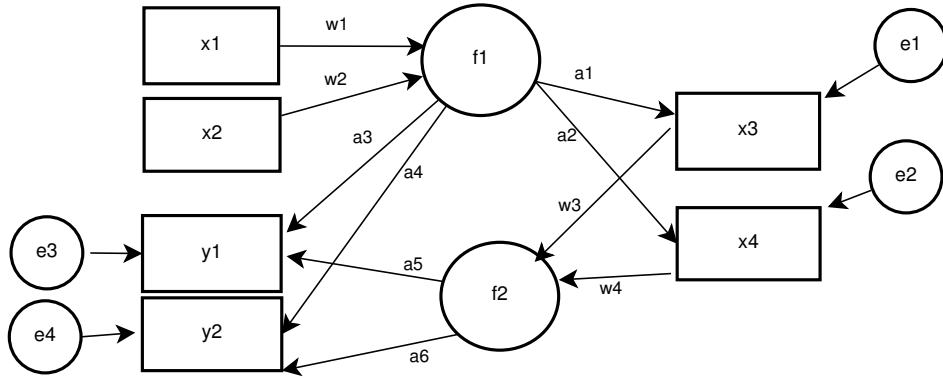


Figure 2: An extensive specification of ERA.

scores $\mathbf{F} = [\mathbf{f}_1|\mathbf{f}_2]$ respectively, and suppose that \mathbf{f}_1 has also an impact on $\mathbf{X}_2 = [\mathbf{x}_3|\mathbf{x}_4]$, the formative block of \mathbf{f}_2 ; this leads to the following structural equations (as depicted in Figure 2):

$$\mathbf{X}_2 = \mathbf{f}_1 \mathbf{a}_1^\top + \mathbf{E}_2, \quad (3)$$

$$\mathbf{Y} = \mathbf{f}_1 \mathbf{a}_2^\top + \mathbf{f}_2 \mathbf{a}_3^\top + \mathbf{E}_1, \quad (4)$$

where $\mathbf{a}_1^\top = (a_1, a_2)$, $\mathbf{a}_2^\top = (a_3, a_4)$, and $\mathbf{a}_3^\top = (a_5, a_6)$ are composite loadings for \mathbf{f}_1 on \mathbf{X}_2 , and for \mathbf{f}_1 and \mathbf{f}_2 on \mathbf{Y} , respectively, whereas $\mathbf{E}_2 = (\mathbf{e}_1|\mathbf{e}_2)$, and $\mathbf{E}_1 = (\mathbf{e}_3|\mathbf{e}_4)$ are two error matrices.

Using the rule that when an observed variable block is exogenous as well as endogenous (\mathbf{X}_2) it is included in both \mathbf{Y} and \mathbf{X} (Takane and Hwang 2005), the associated ERA model becomes

$$\tilde{\mathbf{Y}} = [\mathbf{X}_1|\mathbf{X}_2] \begin{bmatrix} \mathbf{w}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{w}_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & a_5 & a_6 \end{bmatrix} + [\mathbf{E}_2|\mathbf{E}_1] = \mathbf{X} \mathbf{W} \mathbf{A}^\top + \mathbf{E}, \quad (5)$$

where $\tilde{\mathbf{Y}} = [\mathbf{X}_2|\mathbf{Y}] = [\mathbf{x}_3, \mathbf{x}_4|\mathbf{y}_1, \mathbf{y}_2]$ is the new matrix of dependent variables and \mathbf{E} is the new (augmented) matrix of errors.

2.1. Estimation

To estimate parameters, the loss function associated to the model in Equation 1 is

$$SS[\mathbf{Y} - \mathbf{X}\mathbf{W}\mathbf{A}^\top] = SS[\text{vec}(\mathbf{Y}) - \text{vec}(\mathbf{X}\mathbf{W}\mathbf{A}^\top)], \quad (6)$$

where $SS[\mathbf{Z}] = \text{trace}(\mathbf{Z}^\top \mathbf{Z})$ and $\text{vec}(\mathbf{X})$ is the vectorization operator, by which \mathbf{X} columns are stacked one on the other. Notice that \mathbf{W} and \mathbf{A}^\top may contain fixed (zero) elements, depending on the specified model. Unfortunately, unlike statistical methods based on singular value decomposition (SVD) or generalized SVD, minimization of the loss function in Equation 6 cannot be achieved in a closed form, due to the fixed values in \mathbf{W} and \mathbf{A}^\top , containing zeros. Hence, ERA authors employ an alternating least squares (ALS) algorithm, developed by [Kiers and Ten Berge \(1989\)](#). In the algorithm, matrices \mathbf{W} and \mathbf{A}^\top are alternately updated in a two-step algorithm until convergence is reached.

To this end the loss function in Equation 6 may be expressed in two alternative versions

$$SS(\mathbf{Y} - \mathbf{X}\mathbf{W}\mathbf{A}^\top) = SS[\text{vec}(\mathbf{Y}) - (\mathbf{A} \otimes \mathbf{X})\text{vec}(\mathbf{W})] \quad (7)$$

$$= SS[\text{vec}(\mathbf{Y}) - (\mathbf{I} \otimes \mathbf{F})\text{vec}(\mathbf{A}^\top)], \quad (8)$$

where “ \otimes ” is the Kronecker product.

In the first step, \mathbf{W} is updated for fixed \mathbf{A}^\top (using Equation 7) and in the second step, \mathbf{A}^\top is updated for fixed \mathbf{W} (using Equation 8). More explicitly, let \mathbf{w} be the vector obtained by eliminating the zero elements of $\text{vec}(\mathbf{W})$, and let $\mathbf{\Omega}$ be the matrix obtained by eliminating the columns of $\mathbf{A} \otimes \mathbf{X}$ corresponding to the zero elements of $\text{vec}(\mathbf{W})$. The LS estimate of \mathbf{w} therefore is

$$\tilde{\mathbf{w}} = (\mathbf{\Omega}^\top \mathbf{\Omega})^{-1} \mathbf{\Omega}^\top \text{vec}(\mathbf{Y}). \quad (9)$$

The Moore-Penrose generalized inverse is used if $(\mathbf{\Omega}^\top \mathbf{\Omega})$ is singular. \mathbf{W} is then reconstructed from $\tilde{\mathbf{w}}$ and \mathbf{F} is normalized to respect the identification restriction.

In the second step, let \mathbf{a} be the vector obtained by eliminating the zero element of $\text{vec}(\mathbf{A}^\top)$, and let $\mathbf{\Gamma}$ be the matrix obtained by eliminating the columns of $\mathbf{I} \otimes \mathbf{F}$ corresponding to the zero elements of $\text{vec}(\mathbf{A}^\top)$. The LS estimate of \mathbf{a} subsequently is

$$\tilde{\mathbf{a}} = (\mathbf{\Gamma}^\top \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^\top \text{vec}(\mathbf{Y}). \quad (10)$$

The Moore-Penrose generalized inverse is used if $(\mathbf{\Gamma}^\top \mathbf{\Gamma})$ is singular. \mathbf{A}^\top is then reconstructed from \mathbf{a} .

The above steps are iterated until convergence. Nonparametric bootstrap was employed to estimate the standard errors of parameter estimates whereas the critical ratio (CR), i.e., the parameter estimate divided by its standard error, can be used to test the significance of the estimates.

Finally, ERA allows the evaluation of the total fit of a hypothesized model, measured by the total variance of the observed endogenous variables explained by the exogenous variables. The fit index Ψ is

$$\Psi = 1 - \frac{SS[\tilde{\mathbf{Y}} - \mathbf{X}\mathbf{W}\mathbf{A}^\top]}{SS[\tilde{\mathbf{Y}}]}, \quad (11)$$

where $\tilde{\mathbf{Y}}$ collects all endogenous variables. They may be represented by a single block (see Equation 2, where $\tilde{\mathbf{Y}} = \mathbf{Y}$) or by multiple blocks (see Equation 5, where $\tilde{\mathbf{Y}} = [\mathbf{X}_2 | \mathbf{Y}]$). This

fit index ranges from 0 to 1. The larger the fit value, the more variance of the endogenous variables is explained by the exogenous variables.

In accordance with the traditional use of CA, the ERA approach is viewed as purely descriptive, without any pretension of inferential aspects. However, beside the overall fit statistic (Ψ), some other diagnostics and tests on residuals can be used to evaluate the adequacy of the model. For each dependent variable stored in the \mathbf{Y} matrix, residuals can be plotted against the predicted values. There should be no pattern to these plots if residuals are "white-noise". As a formal test we can use the White test for assessing residual homoscedasticity (we suggest this test since it does not require the prior knowledge of the form of heteroscedasticity). Tests for residual randomness, such as a modified version of the Durbin-Watson test and the run test (Pratschke 1971) can also be considered. However, unlike the case of time series data, both tests have limited applicability, since they are affected by subjective choices in ordering the residuals that may give misleading results. More studies are needed in this context.

2.2. Imposing additional constraints on parameters

A variety of structural hypotheses on model parameters (in \mathbf{W} or \mathbf{A}^\top) may be incorporated in the form of linear constraints by the reparameterization method (Takane and Hwang 2005), which specifies the space spanned by column vectors of a constraint matrix \mathbf{H} on a vector of parameters $\boldsymbol{\alpha}$. Let \mathbf{H} define the linear constraints on \mathbf{a} , incorporated by specifying $\mathbf{a} = \mathbf{H}\boldsymbol{\alpha}$, for some vector $\boldsymbol{\alpha}$ of unconstrained parameters. After the estimation of $\boldsymbol{\alpha}$ ($\boldsymbol{\alpha}^*$) obtained by projecting $\Gamma\mathbf{H}$ on $\text{vec}(\mathbf{Y})$, as in Equation 10, the constrained estimate of \mathbf{a} (\mathbf{a}^*) is estimated by $\mathbf{a}^* = \mathbf{H}\boldsymbol{\alpha}^*$. Another possibility to specify constraints (e.g., for \mathbf{a}) is in the null-space form $\mathbf{P}^\top\mathbf{a} = \mathbf{0}$, where \mathbf{P} is a matrix incorporating a system of constraints among parameters (a single constraint is thus incorporated in the vector form $\mathbf{p}^\top\mathbf{a} = 0$). Virtually identical considerations can be made for \mathbf{w} (see Takane and Hwang 2005, for details).

3. The macro %ERA

Excerpts of SAS code will now be given to specify the model of Equation 2 (see Figure 1).

3.1. Path diagram declaration

The structural equations have to be declared in the `cards` of the first part of the macro. `XF` contains the first segment of the path diagram: \mathbf{x}_1 and \mathbf{x}_2 form the LC \mathbf{f}_1 , whereas \mathbf{x}_3 and \mathbf{x}_4 form the LC \mathbf{f}_2 . `FY` contains the second segment of the path diagram: \mathbf{f}_1 goes towards \mathbf{y}_1 and \mathbf{y}_2 , \mathbf{f}_2 does the same.

```
data XF;
  input n v1 $50.;
cards;
1 x01/x02
2 x03/x04
;run;

data FY;
input ne v1e $50.;
```

```
cards;
1 y01/y02
2 y01/y02
;run;
```

3.2. Estimation steps

After manipulating this input into the corresponding variable matrices \mathbf{X} and \mathbf{Y} and parameters matrices \mathbf{W} and \mathbf{A}^\top , the algorithm can start.

Step 1: \mathbf{W} estimate

```
vwd = vectorise(Wd);
Ad = t(ATd);
omegad = Ad @ X;
i = 1;
do while(i <= nrow(vwd));
if vwd[i] ne 0 | wcon[i] = 1 then i = i + 1;
else do;
vwd = delrow(vwd,i); wcon = delrow(wcon,i); omegad = delcol(omegad,i);
end;
end;
vY = vectorise(Y); x
wupd_c = ((ginv(t(B) * t(omegad) * omegad * B)) * t(B) * (t(omegad))) * vY;
wupd = B * wupd_c;
Wd = update(Wd,wupd,wcon2);
Fd = X * Wd; Fd = normalize(Fd);
```

Step 2: \mathbf{A}^\top estimate

```
ad = vectorise(ATd);
Id = I(ncol(ATd));
gammad = Id @ Fd;
i = 1;
do while(i <= nrow(ad));
if ad[i] ne 0 | acon[i]=1 then i = i + 1;
else do;
ad = delrow(ad,i); acon = delrow(acon,i); gammad = delcol(gammad,i);
end;
end;
aupd_c = ((ginv(t(H) * t(gammad) * gammad * H)) * t(H) * (t(gammad))) * vY;
aupd = H * aupd_c;
ATd = update(ATd,aupd,acon2);
```

Estimated parameters for \mathbf{W} and \mathbf{A}^\top (one-shot estimates) are stored in a dataset.

3.3. Bootstrap procedure

The algorithm estimates standard errors using bootstrap, firstly replicating B times (the number of `replic` in the `%ERA` macro) the original dataset.

```
proc surveysselect data=data_nomiss method=urs rate=1 seed=813616000
rep = &replic out = data_boot; run;
/* expands numberhits > 1 */
data data_boot;
  set data_boot;
  do i = 1 to numberhits;
    output;
  end;
  drop i numberhits;
run;
```

then for each of the replications, it proceeds to fit the model as specified above

```
%do rep = 1 %to &replic;
%put EXECUTING REPLICATION &rep OUT OF &replic;
use zdata_boot where (replicate = &rep);
```

The macro uses the multiple datasets sampled with the `surveysselect` procedure. Each of them has an index (`replicate`), which will be matched to the `&rep` macro variable to get the specified bootstrap sample estimates.

3.4. Bootstrap estimates

Using all the stored replications, mean bootstrap estimates, bootstrap s.e. and critical ratios (one-shot parameters divided by their bootstrap s.e.) can be calculated.

Mean bootstrap estimates

```
sumW = J(nrow(W), ncol(W) - 2, 0);
sumA = J(nrow(A), ncol(A) - 2, 0);
%do m = 1 %to &replic;
sumW = sumW + W&m;
sumA = sumA + A&m;
%end;
MEAN_W = sumW / &replic;
MEAN_A = sumA / &replic;
```

Bootstrap standard errors

```
sumsqdiffW = j(nrow(W), ncol(W) - 2, 0);
sumsqdiffA = j(nrow(A), ncol(A) - 2, 0);
%do m = 1 %to &replic;
```

```

sqdiffW&m = (W&m - MEAN_W) \#\# 2;
sqdiffA&m = (A&m - MEAN_A) \#\# 2;
sumsqdiffW = sumsqdiffW + sqdiffW&m;
sumsqdiffA = sumsqdiffA + sqdiffA&m;
%end;
SE_W = (sumsqdiffW / (&replc - 1)) \#\# 0.5;
SE_A = (sumsqdiffA / (&replc - 1)) \#\# 0.5;

```

Bootstrap critical ratios

```

CR_W = j(nrow(W), ncol(W) - 2, 0);
CR_A = j(nrow(A), ncol(A) - 2, 0);
use results_ws;
read all into W_ONESHOT;
use results_as;
read all into A_ONESHOT;
CR_W = (W_ONESHOT / SE_W);
CR_A = (A_ONESHOT / SE_A);

```

At the end of the %ERA macro execution all outputs are stored in the SAS work directory, with the following data tables:

- a_est_oneshot, the one-shot estimates for **W** and **A**.
- b1_est_boot, the bootstrap average estimates for **W** and **A**.
- b2_bias__boot, the bootstrap estimates of bias for **W** and **A**.
- b3_corrected__boot, the bootstrap corrected estimates for **W** and **A**.
- c_se__boot, the bootstrapped standard errors for the estimates.
- d_cr__boot, the bootstrapped critical ratios for the estimates.
- e_psi_oneshot, the fit index Ψ .
- f_replc_w, all the estimates for each bootstrap replicate, for **W**.
- f_replc_a, all the estimates for each bootstrap replicate, for **A**.
- out_era, the output dataset, containing the standardized dataset used for the one-shot estimation with additional variables such as fitted values (endogenous variables) and latent composite scores, as well as residuals and leverages.
- zdata__boot, the data table containing all the bootstrap samples.
- The original data table and the data table with only non-missing observations.

3.5. Convergence and initial values

The ALS steps described in Section 2.1 are iterated in the macro until at least one of the following criteria is met:

- The difference in absolute value between \mathbf{A}^\top (and \mathbf{W}) in the $(k - 1)$ th iteration and in the k -th iteration is lower than the specified threshold c (the value `threshold` in the macro, e.g., $c = 0.0001$).
- The maximum number M of iterations (the value `iter` in the macro) is reached.

```
diffW = abs(Wd - oldWd);
diffAT = abs(ATd - oldATd);
flgw = 1; flgat = 1;
  do i = 1 to nrow(diffW);
    do j = 1 to ncol(diffW);
      if diffW[i,j] > &threshold then flgw = 0;
    end;
  end;
  do i = 1 to nrow(diffAT);
    do j = 1 to ncol(diffAT);
      if diffAT[i,j] > &threshold then flgat = 0;
    end;
  end;
ciclo = ciclo - 1;
if ciclo = 0 | (flgw = 1 & flgat = 1) then mustexit = 1;
```

To ensure faster convergence and to minimize the risk of local minima, as suggested by [Takane and Hwang \(2005\)](#), starting values for \mathbf{W} and \mathbf{A}^\top are chosen as follows:

- The PCA solutions on the \mathbf{X} matrix are computed and used as starting values for \mathbf{W} , and the principal coordinates are used to obtain \mathbf{F} .
- Given \mathbf{F} , LS estimates of \mathbf{F} on \mathbf{Y} are used as initial values for \mathbf{A}^\top .

Initial values for \mathbf{W}

```
call eigen(eigval, eigvec, t(L&i) * L&i);
feigvec = eigvec[, 1];
vec&i = j(ncol(L&i), 1, 1);
if (&startw = 0 & &starta = 0) then vec&i = feigvec;
else vec&i = vec&i * &startw;
```

Initial values for \mathbf{A}^\top

```
AStart = inv(t(F) * F) * t(F) * Y;
if (&startw = 0 & &starta = 0) then
ATd = AT \# AStart;
else ATd = AT * (&starta);
```

3.6. Imposition of linear constraints

Restrictions on parameters can be applied, as shown in Section 2.2, by declaring the matrices of linear constraints \mathbf{B} and \mathbf{H} (for \mathbf{W} and \mathbf{A}^\top respectively), for example by imposing the

constraint $w_4 = 0$ (and no constraint on \mathbf{A}^\top) in the path of Figure 1. This leads to the following constraint matrices:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The code below specifies \mathbf{B} and \mathbf{H} .

Constraints $\mathbf{w} = \mathbf{B}\mathbf{v}$, \mathbf{B} matrix declaration

```
data con_w;
  input w01-w04;
cards;
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 0
;run;
```

Constraints $\mathbf{a} = \mathbf{H}\mathbf{c}$, \mathbf{H} matrix declaration

```
data con_a;
  input a01-a04;
cards;
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
;run;
```

The estimation procedure thus incorporates these two matrices into the steps of the algorithm.

\mathbf{w} estimate, with \mathbf{B} matrix embedded

```
wupd_c = ((ginv(t(B) * t(omegad) * omegad * B)) * t(B) * (t(omegad))) * vY;
wupd = B * wupd_c;
```

\mathbf{a} estimate, with \mathbf{H} matrix embedded

```
aupd_c = ((ginv(t(H) * t(gammad) * gammad * H)) * t(H) * (t(gammad))) * vY;
aupd = H * aupd_c;
```

3.7. Macro invocation

The code below invokes the macro.

```
%macro ERA(dataset, iter, replic, threshold, startw, starta);
```

The macro parameters are:

- **dataset**: The input SAS dataset to be used for the model, including the library path (e.g., `lib.data` with `lib` as the SAS library).
- **iter**: The maximum number of iterations M of the ALS algorithm. Must be integer and greater than 1 (e.g., $M = 100$). This parameter is essential to stop the macro if convergence is not reached in M ALS iterations. An output alert message will be shown in the SAS log window if convergence does not occur.
- **replic**: The number of Bootstrap replications to be generated to obtain standard errors and critical ratios for the estimates. Must be integer and greater than 1 (e.g., $B = 100$).
- **threshold**: The convergence threshold c for the estimates. Must be between 0 and 1 (e.g., $c = 0.0001$). This parameter has higher priority than **iter**: If convergence is reached before the maximum number M of ALS iterations the algorithm can stop. However, *both* parameters must be properly declared in the macro.
- **startw** and **starta**: Optionally inputted starting values for \mathbf{W} and \mathbf{A}^\top . If *both* values are zero, PCA and LS are performed to obtain initial values for the matrices. If one of the two values is zero, but not the other, the macro exits with error. If both values are numerical and not zero, they are chosen as initial estimates *for all the non-zero values of the matrices*. For example, if **startw** = 0.3 and **starta** = 0.7, given the model in Figure 1, the starting matrices will be

$$\mathbf{W} = \begin{bmatrix} 0.3 & 0 \\ 0.3 & 0 \\ 0 & 0.3 \\ 0 & 0.3 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0.7 & 0.7 \\ 0.7 & 0.7 \end{bmatrix}.$$

3.8. User-defined functions

To successfully implement the ERA algorithm, several functions have been used in the script:

- **delrow(x, i)**, that deletes the i -th row from a generic matrix \mathbf{X} .
- **delcol(x, i)**, that deletes the i -th column from a generic matrix \mathbf{X} .
- **vectorise(x)**, that applies the vectorization operator to a generic matrix \mathbf{X} .
- **normalize(x)**, that transforms a generic matrix \mathbf{X} so that $\text{diag}(\mathbf{X}^\top \mathbf{X}) = \mathbf{I}$.
- **update(X, y, c)**, that reconstructs the original matrix \mathbf{X} to which row/column elimination has been previously applied.

4. Application of the macro %ERA

To assess the %ERA macro performance, it will be used in two empirical studies that the authors performed in the paper where the ERA model was proposed. An attempt to replicate the same

results will be made, firstly in an empirical application using the World Health Organization (WHO) data, and secondly in a simulation study where the authors analyzed the small sample behavior of ERA.

4.1. The WHO data application

As far as the empirical study is concerned, Takane and Hwang (2005) selected 51 nations from the 1999 World Health Report, published by the World Health Organization in the United Nations, for which the data were non-missing for the following variables: infant mortality rate (IMR), maternal mortality ratio (MMR), – both considered as endogenous variables – and the following as exogenous variables: real gross domestic product (GDP) per capita adjusted for purchasing power parity, expressed in 1985 US dollars, the average number of years of education given for females aged 25 years and above (FEDU), the percentage of children immunized against measles in 1997 (Measles), and total health expenditures as a percentage of GDP in 1995 (HealthExp). The authors identify two components for the last four observed exogenous variables. One component called "social and economic factor" (SE) was defined as a linear combination of GDP and FEDU, and the other called "health services factor" (HS) composed of Measles and HealthExp. The specified two-component model (Equation 2) is the same of that presented in Figure 1, where $\mathbf{X}_1 = [\text{GDP}, \text{FEDU}]$, $\mathbf{X}_2 = [\text{Measles}, \text{HealthExp}]$, $\mathbf{Y} = [\text{IMR}, \text{MMR}]$ and $\mathbf{F} = [\text{SE}, \text{HS}]$. With the %ERA macro this model was fitted to the data, using the same specification as in the original paper, such as to use 100 bootstrap samples to estimate standard errors and critical ratios.

```
data XF;
  input n v1 $50.;
cards;
1 GDP/FEDU
2 measles/HEXP
;run;
```

```
data FY;
input ne vle $50.;
cards;
1 IMR/MMR
2 IMR/MMR
;run;
```

```
%ERA(dataset = who, iter = 100, replic = 100, threshold =
  0.0001, startw = 0, starta = 0);
```

Table 1 reports the original ERA results (thereby called “H-T”, the initials of the authors) and those reproduced by the macro %ERA, for the unconstrained model specified in Figure 1, as well as for the constrained model. In the latter we assumed the component weight for Healthexp to be zero ($w_4 = 0$) by imposing the constraint matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Path coefficient	Unconstrained		Constrained	
	Est. (S.E.)	Est. (S.E.)	Est. (S.E.)	Est. (S.E.)
	%ERA	H-T	%ERA	H-T
GDP → SE (w_1)	0.50 (0.18)	-0.50 (0.17)	0.49 (0.17)	-0.49 (0.16)
FEDU → SE (w_2)	0.58 (0.17)	-0.57 (0.16)	0.58 (0.16)	-0.57 (0.16)
Measles → HS (w_3)	0.98 (0.16)	-0.96 (0.12)	1.01 (0.00)	-1.00 (0.00)
HealthExp → HS (w_4)	0.12 (0.22)	-0.13 (0.24)	0.00	0.00
SE → IMR (a_1)	-0.58 (0.11)	0.58 (0.10)	-0.61 (0.10)	0.61 (0.11)
SE → MMR (a_2)	-0.44 (0.11)	0.43 (0.10)	-0.47 (0.10)	0.47 (0.13)
HS → IMR (a_3)	-0.42 (0.12)	0.41 (0.09)	-0.40 (0.11)	0.40 (0.11)
HS → MMR (a_4)	-0.45 (0.14)	0.45 (0.11)	-0.44 (0.13)	0.43 (0.15)
FIT (Ψ)	0.6597	0.6512	0.6579	0.6491

Table 1: Parameters estimates (Est.) and standard errors (S.E.) of the macro ERA (%ERA) and those reported in the original ERA paper (H-T) for the WHO dataset.

Path coefficient	Unconstrained estimates		Constrained estimates	
	Mean bootstrap	Bias corrected	Mean bootstrap	Bias corrected
GDP → SE (w_1)	0.50	0.49	0.48	0.50
FEDU → SE (w_2)	0.57	0.58	0.58	0.58
Measles → HS (w_3)	0.94	1.01	1.01	1.01
HealthExp → HS (w_4)	0.14	0.10	0.00	0.00
SE → IMR (a_1)	-0.60	-0.55	-0.63	-0.58
SE → MMR (a_2)	-0.47	-0.41	-0.50	-0.44
HS → IMR (a_3)	-0.40	-0.43	-0.38	-0.43
HS → MMR (a_4)	-0.43	-0.47	-0.41	-0.46

Table 2: Mean bootstrap estimates and bias corrected estimates of the macro ERA, for the WHO dataset.

In Table 1 bootstrapped standard errors are obtained with 100 bootstrap samples, for both the unconstrained and constrained model. To fit the constrained model, before the macro invocation of the unconstrained model (and after the input of both path diagrams) the following additional code must be inserted:

```
data con_w;
input w01-w04;
cards;
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 0
;run;

%ERA(dataset = who, iter = 100, replic = 100, threshold = 0.0001,
startw = 0, starta = 0);
```

Differences between the estimates are quite negligible and are probably due to technical details

in the macro such as convergence thresholds or different numerical methods for approximation used by the authors. Moreover, as additional results, Table 2 reports the mean of the bootstrap estimates and the “bias corrected” versions of estimated parameters.

For what concerns diagnostics in the unconstrained model, Figure 3 reports, for each dependent variable, the plot between residuals and predicted values. Homoscedasticity is clearly rejected at a 0.05 significance level, since $p = 0.02$ for IMR and $p = 0.03$ for MMR. The result is also confirmed by Figure 3.

Finally, to complete the picture of diagnostics, Figure 4 reporting, for each dependent variable, studentized residuals versus leverages, does not suggest any influential observation, but only few outliers and unusual observations.

4.2. The simulation study

The same simulation study which appeared in the original paper of Takane and Hwang (2005) is conducted with the macro %ERA. Specifically the model in Equation 2, detailed in Figure 1, is chosen with the following parameter values:

$$\mathbf{W} = \begin{bmatrix} 0.6 & 0 \\ 0.6 & 0 \\ 0 & 0.6 \\ 0 & 0.6 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}.$$

For the generation of \mathbf{X} and \mathbf{Y} , normality assumption was taken, with $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Theta)$ and $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where Θ and Σ are

$$\Theta = \begin{bmatrix} 1 & & & \\ 0.3 & 1 & & \\ 0.1 & 0.1 & 1 & \\ 0.1 & 0.1 & 0.3 & 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \\ 0.1 & 1 \end{bmatrix}.$$

The sample sizes considered for this Monte Carlo simulation consisting of $N = 1000$ samples are $n = 100, 200, 400$, with \mathbf{X} generated only once for each sample size. Furthermore, the initial values of the parameters used in the algorithm are indeed the true values, to ensure a faster convergence to the global minimum.

To measure the degree of recovery of the population parameters the mean congruence coefficient between real parameters (θ) and estimates ($\hat{\theta}$) stacked vectors is used

$$\phi = \frac{\theta^\top \hat{\theta}}{\sqrt{(\theta^\top \theta)(\hat{\theta}^\top \hat{\theta})}}.$$

For the simulation the %ERA macro has been slightly modified in terms of input handling and output datasets. Details are not provided since it is out of the scope of this paper.

A deep look at the results of Table 3 also indicates two main considerations:

- When $n \geq 200$ the %ERA loadings as a group appear less biased than those of H-T, (with unclear pattern when $n < 200$) whereas standard errors are always higher than those of H-T.

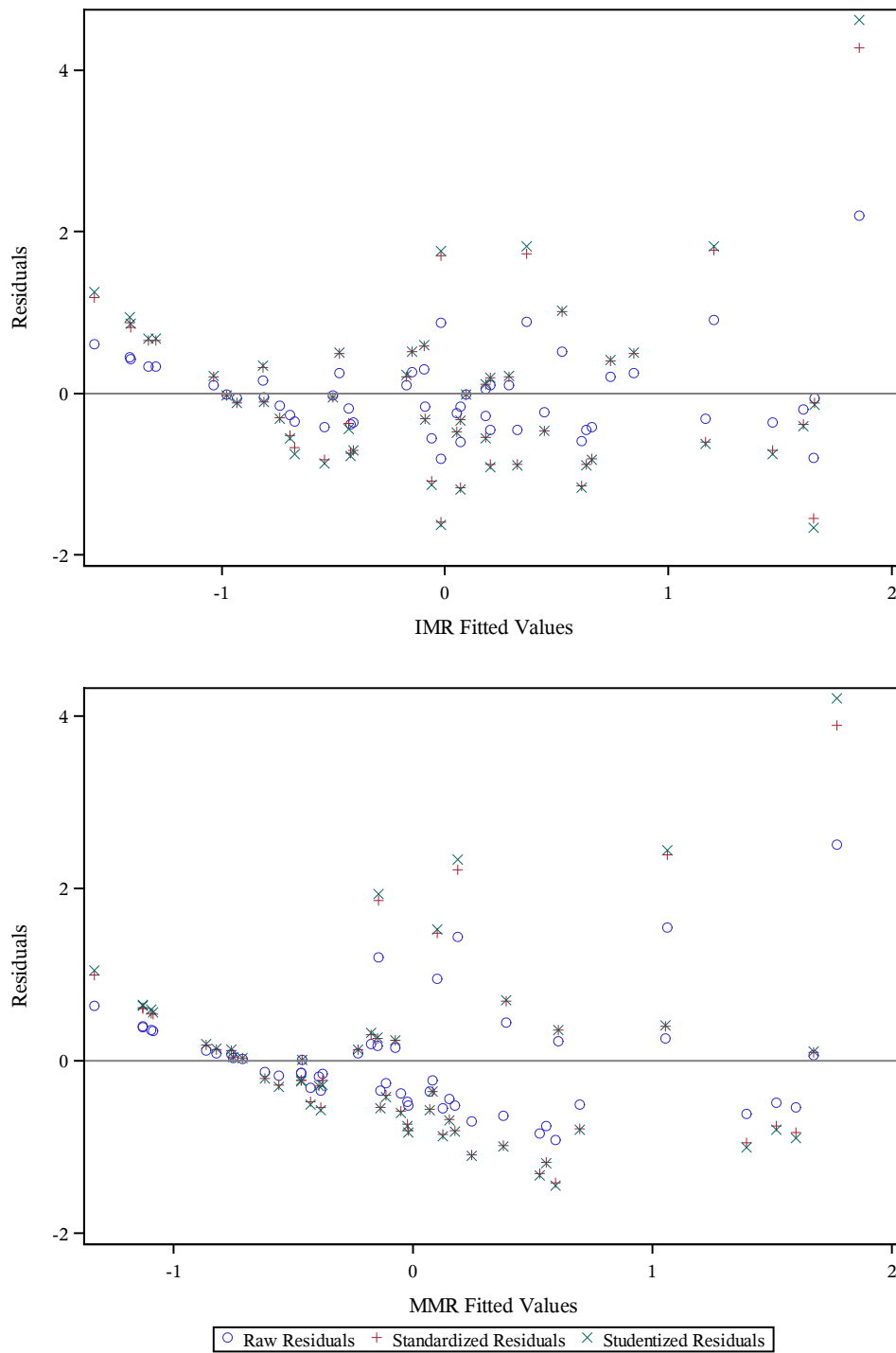


Figure 3: Residuals (raw, standardized and studentized) against fitted values, for each dependent variable.

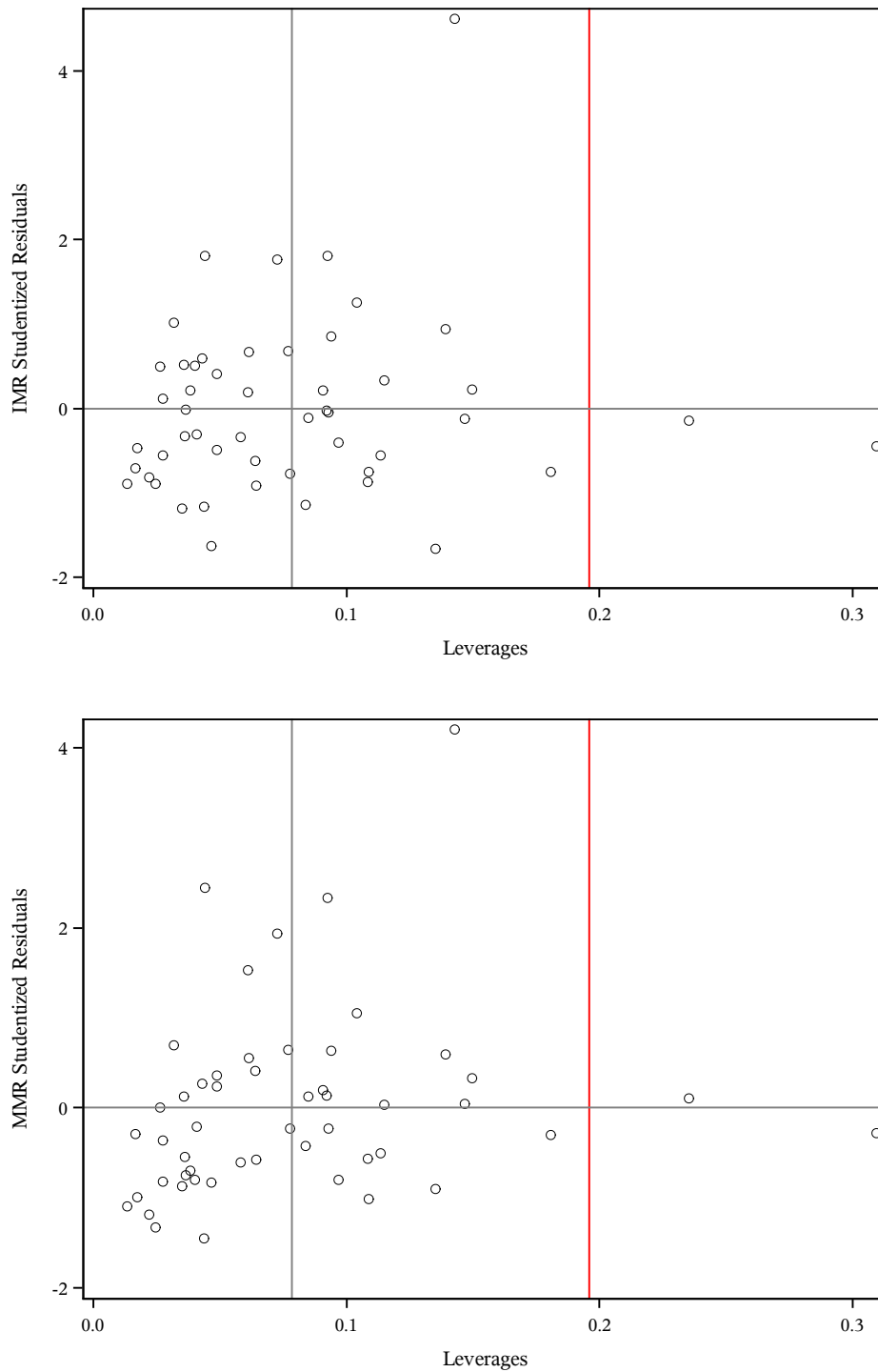


Figure 4: Studentized residuals against leverages, for each dependent variable. The red vertical line represents the threshold for unusual observations.

$n = 100$		%ERA macro		H-T results	
Par.	Value	Est.	S.E.	Est.	S.E.
w_1	0.6	0.576	0.364	0.592	0.335
w_2	0.6	0.571	0.380	0.558	0.341
w_3	0.6	0.598	0.362	0.586	0.319
w_4	0.6	0.570	0.376	0.550	0.334
a_1	0.2	0.199	0.111	0.200	0.102
a_2	0.2	0.193	0.109	0.214	0.101
a_3	0.2	0.193	0.109	0.206	0.100
a_4	0.2	0.190	0.108	0.219	0.099
Congruence		$\phi = 0.83$		$\phi = 0.87$	

$n = 200$		%ERA macro		H-T results	
Par.	Value	Est.	S.E.	Est.	S.E.
w_1	0.6	0.606	0.238	0.629	0.255
w_2	0.6	0.622	0.248	0.561	0.265
w_3	0.6	0.602	0.247	0.592	0.255
w_4	0.6	0.611	0.230	0.587	0.265
a_1	0.2	0.193	0.072	0.181	0.072
a_2	0.2	0.189	0.075	0.184	0.069
a_3	0.2	0.194	0.072	0.193	0.069
a_4	0.2	0.195	0.074	0.192	0.069
Congruence		$\phi = 0.94$		$\phi = 0.92$	

$n = 400$		%ERA macro		H-T results	
Par.	Value	Est.	S.E.	Est.	S.E.
w_1	0.6	0.593	0.169	0.596	0.169
w_2	0.6	0.614	0.164	0.605	0.170
w_3	0.6	0.605	0.169	0.630	0.162
w_4	0.6	0.608	0.166	0.566	0.169
a_1	0.2	0.198	0.051	0.189	0.047
a_2	0.2	0.198	0.051	0.194	0.047
a_3	0.2	0.197	0.051	0.193	0.048
a_4	0.2	0.197	0.052	0.200	0.047
Congruence		$\phi = 0.97$		$\phi = 0.96$	

Table 3: True parameter (Value), estimate (Est.), standard error (S.E.) and congruence coefficient by sample size (n) for the macro ERA (%ERA macro) and those reported in the original ERA paper (H-T results).

- When $n \geq 200$, the %ERA weights standard errors are generally lower than those of H-T, whereas there is not a clear pattern concerning the bias.

Trying to explain this trade-off between bias and standard error for loading (at least for the

%ERA macro), simulation results for sufficiently large samples confirm a direct relation between reduction of biases (at $2n$ with respect to n) and reduction of standard errors: Components loadings, exhibiting lower bias reduction, also demonstrate a lower reduction of standard errors. Moreover, the fact that at different n the best approach (%ERA and H-T) in term of bias does not outperform the other in term of standard errors may reflect that ERA, and ALS algorithms in general, are prone to the problem of local minima, especially for loading parameters.

5. Conclusion

Extended redundancy analysis is a data-analytical technique for complex structural models in a component analysis framework. The ERA estimates component weights and component loadings by minimizing an overall model fit, such as the sum of squares of discrepancies between the observed endogenous variables and their predicted counterparts from the exogenous without any explicit distributional assumptions. Although ERA has potential application in diverse areas within and outside the behavioral sciences, its use is hampered by a main issue: Easily accessible software for estimating ERA models is lacking so far. In this paper, we address this issue, presenting the %ERA macro, which was built in SAS. Two possible extensions can be proposed in future works: The first is the inclusion of external covariates which do not belong to formative LC indicators, but instead may have an impact on these LCs and a direct effect on the endogenous variables too; the second extension is linked to a specification of structural links between an exogenous and an endogenous LC that, at present, must be mediated by the formative indicators of an endogenous LC. Potential candidates to accommodate both issues may be found in more general methods, such as Generalized redundancy analysis, which may offer promising results (Lovaglio and Vittadini 2014).

References

- Bougeard S, Qannari EM, Lupo C, Hanafi M (2011). “From Multiblock Partial Least Squares to Multiblock Redundancy Analysis. A Continuum Approach.” *Informatica*, **22**(1), 11–26.
- Davies PT, Tso MKS (1982). “Procedures for Reduced-Rank Regression.” *Journal of the Royal Statistical Society C*, **31**(3), 244–255. doi:10.2307/2347998.
- Kiers HAL, Ten Berge JMF (1989). “Alternating Least Squares Algorithms for Simultaneous Components Analysis with Equal Component Weight Matrices in Two or More Populations.” *Psychometrika*, **54**(3), 467–473. doi:10.1007/bf02294629.
- Lovaglio PG, Vittadini G (2014). “Structural Equation Models in a Redundancy Analysis Framework with Covariates.” *Multivariate Behavioral Research*, **49**(5), 486–501. doi:10.1080/00273171.2014.931798.
- Meredith W, Millsap RE (1985). “On Component Analyses.” *Psychometrika*, **50**(4), 495–507. doi:10.1007/bf02296266.
- Pratschke JL (1971). “On Testing for Uniformity of Fit in Regression: An Econometric Case Study.” *Economic and Social Review*, **2**(4), 497–505.

- Reinsel GC, Velu RP (1998). *Multivariate Reduced-Rank Regression*. Springer-Verlag. doi:
[10.1007/978-1-4757-2853-8](https://doi.org/10.1007/978-1-4757-2853-8).
- SAS Institute Inc (2011). *SAS/STAT Software, Version 9.3*. Cary, NC. URL <http://www.sas.com/>.
- Schonemann PH, Steiger JH (1976). “Regression Component Analysis.” *British Journal of Mathematical and Statistical Psychology*, **29**(2), 175–189. doi:[10.1111/j.2044-8317.1976.tb00713.x](https://doi.org/10.1111/j.2044-8317.1976.tb00713.x).
- Takane Y, Hwang H (2005). “An Extended Redundancy Analysis and Its Applications to Two Practical Examples.” *Computational Statistics & Data Analysis*, **49**(3), 785–808. doi:
[10.1016/j.csda.2004.06.004](https://doi.org/10.1016/j.csda.2004.06.004).
- Van den Wollenberg AL (1977). “Redundancy Analysis: An Alternative to Canonical Correlation Analysis.” *Psychometrika*, **42**(2), 207–219. doi:[10.1007/bf02294050](https://doi.org/10.1007/bf02294050).

Affiliation:

Pietro Giorgio Lovaglio, Gianmarco Vacca
Department of Statistics and Quantitative Methods
University of Milano-Bicocca
20161 Milan, Italy
E-mail: piergioorgio.lovaglio@unimib.it, g.vacca@campus.unimib.it