

# On the relationships among latent variables and residuals in PLS path modeling: The formative-reflective scheme

Giorgio Vittadini, Simona C. Minotti\*, Marco Fattore, Pietro G. Lovaglio

*Dipartimento di Statistica, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy*

Received 4 December 2005; received in revised form 26 October 2006; accepted 26 October 2006

Available online 27 November 2006

## Abstract

A new approach for the estimation and the validation of a structural equation model with a formative-reflective scheme is presented. The basis of the paper is a proposal for overcoming a potential deficiency of PLS path modeling. In the PLS approach the reflective scheme assumed for the endogenous latent variables (LVs) is inverted; moreover, the model errors are not explicitly taken into account for the estimation of the endogenous LVs. The proposed approach utilizes all the relevant information in the formative manifest variables (MVs) providing solutions which respect the causal structure of the model. The estimation procedure is based on the optimization of the redundancy criterion. The new approach, entitled redundancy analysis approach to path modeling (RA-PM) is compared with both traditional PLS Path Modeling and LISREL methodology, on the basis of real and simulated data.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Latent variables; Partial least squares; PLS path modeling; Redundancy analysis; LISREL model

## 1. Introduction

This paper proposes a new approach for the estimation and the validation of a structural equation model characterized by a formative scheme for exogenous latent variables (LVs) and a reflective scheme for endogenous LVs. The proposal consists of an alternative approach to the PLS path modeling approach (PLS-PM) introduced by Lohmöller (1989).

In order to examine the implications involved in the adoption of a formative-reflective scheme, let us examine an example regarding the causal links between the investment in human capital (HC), (i.e. educational attainment and work experience), and the ability to generate earned and property income, as indicated in Fig. 1.

The HC model, which is described in detail in Section 7, consists of two exogenous LVs, “educational HC” and “work experience HC”, and two endogenous LVs which can be defined as the “ability to generate earned income” and the “ability to generate property income”. These two endogenous LVs represent the innate capability to transform educational attainment and work experience into the process of accumulation of earned and property income.

The results of the application reveal some critical points in PLS-PM. First of all, the goodness of fit is quite low ( $GoF = 0.3863$ ) and, as Goldstein–Dillon’s  $\rho$  shows, the formative block containing the Job indicators is not unidimensional ( $\rho = 0.1630$ ). We expect that the extraction of further information from the formative manifest variables (MVs) will improve the model’s goodness of fit. Hence we suggest the explicit introduction of the model errors (estimated

\* Corresponding author. Tel.: +39 2 64485836; fax: +39 2 64485899.

E-mail address: [simona.minotti@unimib.it](mailto:simona.minotti@unimib.it) (S.C. Minotti).

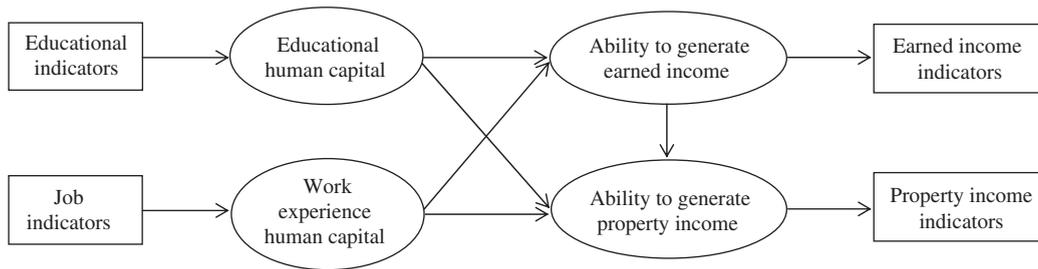


Fig. 1. The structural model for human capital.

by means of the residuals) into PLS-PM. Secondly, we observe that the redundancy of the reflective MVs accounted for by the estimated endogenous LVs (40.62%) is greater than the redundancy accounted for by the formative MVs (34.51%). This anomalous result depends on Lohmöller’s algorithm, which estimates the endogenous LVs as linear combinations of their reflective MVs.

In order to resolve the problems mentioned above, we propose a new approach based on the optimization of the redundancy criterion (Stewart and Love, 1968). The new approach, entitled RA-PM (redundancy analysis approach to path modeling), utilizes all the relevant information in the formative MVs and provides solutions which respect the causal structure of the model.

The paper is organized as follows: in Section 2, the formative-reflective model is introduced; in Section 3, the traditional methodologies utilized for its estimation are discussed; in Section 4, the RA-PM algorithm is presented; in Section 5, an appropriate measure of goodness of fit is proposed; in Section 6, a discussion of the basic choices and properties of RA-PM is provided; in Sections 7 and 8, RA-PM is compared with traditional PLS-PM and the LISREL methodology, on the basis of real and simulated data; in Section 9, some conclusions and further research are proposed.

## 2. The formative-reflective model

The non-recursive structural equation model considered in this paper is characterized by a formative scheme for the exogenous LVs and a reflective scheme for the endogenous LVs.

The *measurement model* of the exogenous LVs imposes a formative relationship between each LV and its corresponding MVs, i.e. each exogenous LV summarizes its MVs

$$\xi_{(j)} = \omega'_{(j)} \underline{X}_{(j)} + u_{(j)}, \quad j = 1, \dots, r, \tag{1}$$

where  $\xi_{(j)}$ ,  $j = 1, \dots, r$ , is a random variable denoting the  $j$ th exogenous LV;  $\underline{X}_{(j)} = (X_{(j)1}, \dots, X_{(j)p_j})'$  is a  $p_j$ -dimensional vector of centred observable random variables;  $\omega_{(j)} = (\omega_{(j)1}, \dots, \omega_{(j)p_j})'$  is a  $p_j$ -dimensional vector of unknown outer weights;  $u_{(j)}$  is a random disturbance, with expected value null and uncorrelated to the MVs  $\underline{X}_{(j)}$ . It can be observed that, in the formative scheme, the blocks of MVs can be multidimensional (Tenenhaus et al., 2005).

The *measurement model* of the endogenous LVs imposes a reflective relationship between each LV and its corresponding MVs, i.e. each endogenous LV is described by its MVs:

$$\underline{Y}_{(k)} = \lambda_{(k)} \eta_{(k)} + \varepsilon_{(k)}, \quad k = 1, \dots, s, \tag{2}$$

where  $\eta_{(k)}$ ,  $k = 1, \dots, s$ , is a random variable which denotes the  $k$ th endogenous LV;  $\underline{Y}_{(k)} = (Y_{(k)1}, \dots, Y_{(k)q_k})'$  is a  $q_k$ -dimensional vector of centred observable random variables;  $\lambda_{(k)} = (\lambda_{(k)1}, \dots, \lambda_{(k)q_k})'$  is a  $q_k$ -dimensional vector of unknown loadings;  $\varepsilon_{(k)} = (\varepsilon_{(k)1}, \dots, \varepsilon_{(k)q_k})'$  is a  $q_k$ -dimensional vector of errors of measurement, with expected value null and uncorrelated to  $\eta_{(k)}$ . It can be observed that, in the reflective scheme, the blocks of MVs are unidimensional in accordance with factor analysis (Tenenhaus et al., 2005).

The model is completed by the *structural equation* which describes the relationship between the LVs

$$\eta_{(k)} = \beta'_{(k)} \underline{\eta} + \gamma'_{(k)} \underline{\xi} + \zeta_{(k)}, \quad k = 1, \dots, s, \tag{3}$$

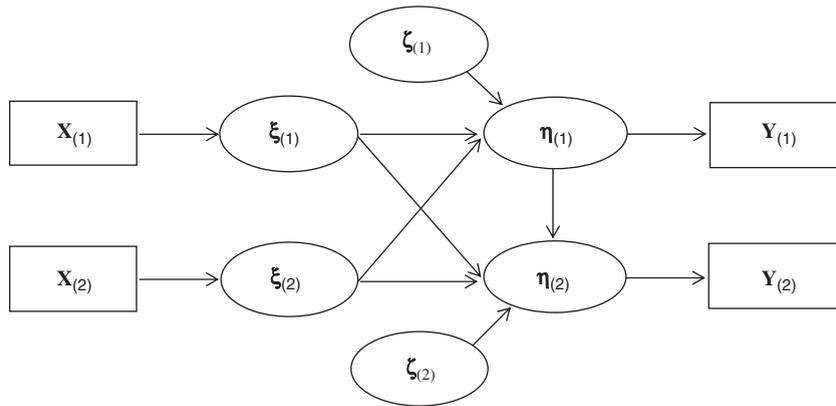


Fig. 2. A formative-reflective structural equation model.

where  $\underline{\eta}$  is a random vector which contains the endogenous LVs;  $\underline{\xi}$  is a random vector which contains the exogenous LVs;  $\underline{\beta}^{(k)}$  and  $\underline{\gamma}^{(k)}$  are vectors of  $s$  and  $r$  unknown path coefficients, respectively;  $\zeta_{(k)}$  is a random variable indicating the “error in equation” associated with the  $k$ th endogenous LV  $\eta_{(k)}$ . Errors in equations “may represent the effect of unknown variables or the effect of known but omitted variables” (Saris and Stronkhorst, 1984).

In a compact form, model (1)–(3) can be rewritten as

$$\underline{\xi} = \underline{\Omega}'_X \underline{X} + \underline{u}, \quad (4)$$

$$\underline{Y} = \underline{\Lambda}'_Y \underline{\eta} + \underline{\varepsilon} \quad (5)$$

and

$$\underline{\eta} = \underline{B}' \underline{\eta} + \underline{\Gamma}' \underline{\xi} + \underline{z} = (\mathbf{I} - \underline{B}')^{-1} \underline{\Gamma}' \underline{\xi} + (\mathbf{I} - \underline{B}')^{-1} \underline{z}. \quad (6)$$

Given the observations of  $n$  independent drawings of the random vectors  $\underline{x}_{(j)}$  and  $\underline{y}_{(k)}$ , by arranging the observations in  $n$ -dimensional column vectors, model (4)–(6) becomes

$$\underline{\Xi} = \underline{X} \underline{\Omega}_X + \underline{U}, \quad (7)$$

$$\underline{Y} = \underline{H} \underline{\Lambda}_Y + \underline{E} \quad (8)$$

and

$$\underline{H} = \underline{H} \underline{B} + \underline{\Xi} \underline{\Gamma} + \underline{Z} = \underline{\Xi} \underline{\Gamma} (\mathbf{I} - \underline{B})^{-1} + \underline{Z} (\mathbf{I} - \underline{B})^{-1}, \quad (9)$$

where  $\underline{\Xi} = (\xi_{(1)}, \dots, \xi_{(r)})$  is a  $(n \times r)$  matrix of latent scores on  $\underline{\xi}$ ;  $\underline{H} = (\eta_{(1)}, \dots, \eta_{(s)})$  is a  $(n \times s)$  matrix of latent scores on  $\underline{\eta}$ ;  $\underline{X} = (\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(p)})$  is a  $(n \times p)$  matrix of observations on  $\underline{X}$ ;  $\underline{Y} = (\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)})$  is a  $(n \times q)$  matrix of observations on  $\underline{Y}$ ;  $\underline{U}$  is a  $(n \times r)$  matrix of random disturbances;  $\underline{E} = (\epsilon_{(1)}, \dots, \epsilon_{(q)})$  is a  $(n \times q)$  matrix of errors of measurement;  $\underline{Z} = (\zeta_{(1)}, \dots, \zeta_{(s)})$  is a  $(n \times s)$  matrix of errors in equations.

In the special case of  $r = 2$  and  $s = 2$ , model (7)–(9) corresponds to the HC model, completed by the errors in equations (see Fig. 2).

A suitable estimation method for model (7)–(9) should have the following properties:

- (i) The parameter identifiability.
- (ii) The non-correlation between exogenous LVs and errors in equations.
- (iii) The uniqueness of the latent scores.
- (iv) The coherence of the solutions with the causal structure of the model (i.e. the path directions).
- (v) The utilization of all the relevant information in the formative MVs, which implies the modelization of the residuals.

### 3. Problems in PLS path modeling and LISREL model

PLS-PM (Wold, 1982) satisfies properties (i) and (iii) (given that “the rank of components models is equal to the rank of the MVs”, Lohmöller, 1989) but does not respect properties (ii), (iv) and (v), as will be discussed in the following section. For the sake of simplicity, the discussion will regard the case of a formative-reflective model characterized by  $r = 1$  and  $s = 1$ .

In traditional PLS-PM, LVs are estimated as linear combinations of their MVs, with weights obtained by means of an iterative procedure which takes into account the inner relations of the model, as will now be briefly described (for an up-to-date review, see Tenenhaus et al., 2005).

*Outer estimation:* The standardized “outer estimates” of latent score vectors  $\xi$  and  $\eta$ , indicated by  $\hat{\xi}_{\text{outer}}$  and  $\hat{\eta}_{\text{outer}}$ , respectively, are obtained as linear combinations of their observed MVs, given the arbitrary vectors of outer weights  $\omega_0$  and  $\nu_0$ :

$$\hat{\xi}_{\text{outer}} \propto \mathbf{X}\omega_0 \quad (10)$$

and

$$\hat{\eta}_{\text{outer}} \propto \mathbf{Y}\nu_0. \quad (11)$$

*Inner estimation:* The standardized “inner estimates” of  $\xi$  and  $\eta$ , indicated by  $\hat{\xi}_{\text{inner}}$  and  $\hat{\eta}_{\text{inner}}$ , respectively, are obtained as linear combinations of standardized “outer estimates” of their adjacent LVs

$$\hat{\xi}_{\text{inner}} \propto \gamma\hat{\eta}_{\text{outer}} \quad (12)$$

and

$$\hat{\eta}_{\text{inner}} \propto \beta\hat{\xi}_{\text{outer}}, \quad (13)$$

where the inner weights  $\hat{\gamma}$  and  $\hat{\beta}$  can be alternatively defined by means of the centroid scheme, the factorial scheme or the path weighting scheme.

*Estimation modes for the outer weights:* The outer weights  $\omega_0$  are updated by means of Mode B (i.e. a multiple regression of the “inner estimate”  $\hat{\xi}_{\text{inner}}$  on matrix  $\mathbf{X}$ ), while the weights  $\nu_0$  are updated by means of Mode A (simple regressions of the vectors  $\mathbf{y}_h$ ,  $h = 1, \dots, q$ , on the “inner estimate”  $\hat{\eta}_{\text{inner}}$ )

$$\hat{\omega} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\xi}_{\text{inner}} \quad (14)$$

and

$$\hat{\nu}' = (\hat{\eta}'_{\text{inner}}\hat{\eta}_{\text{inner}})^{-1}\hat{\eta}'_{\text{inner}}\mathbf{Y} = \hat{\eta}'_{\text{inner}}\mathbf{Y}. \quad (15)$$

The procedure is iterated until the algorithm converges, producing

$$\xi \propto \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\eta} \quad (16)$$

and

$$\hat{\eta} \propto \mathbf{Y}\mathbf{Y}'\xi. \quad (17)$$

From (16) and (17) it follows that PLS-PM does not respect properties (iv) and (v) given that:

- The final estimate  $\xi$  in (16) is proportional to the final estimate  $\hat{\eta}$  and this does not respect the causal relation implied in (3). Moreover, the final estimate  $\hat{\eta}$  in (17) is a linear combination of the columns of matrix  $\mathbf{Y}$  and this does not respect the reflective scheme assumed in (5), although the outer weights, by means of Mode A, preserve this assumption.
- The inner estimate  $\hat{\eta}_{\text{inner}}$  in (13) does not explicitly take into account the “error in equation”, i.e. further components underlied by the block of formative MVs, which can be multidimensional (Tenenhaus et al., 2005).

It can be observed that the LISREL model (Jöreskog, 1973, 1981, 2000) satisfies properties (ii), (iv) and (v), but the formative scheme assumed in (1) is substituted by a reflective scheme, which implies that properties (i) and (iii) are not respected.

In fact, in the LISREL model instead of (7) we have

$$\mathbf{X} = \Xi\Lambda_X + \Delta \quad (18)$$

and consequently the reduced form of model (7)–(9) is

$$\mathbf{J} = \mathbf{M}\mathbf{L}_M + \mathbf{T}, \quad (19)$$

with a covariance structure equal to

$$\Sigma_J = \mathbf{L}'_M \Sigma_M \mathbf{L}_M + \Sigma_T, \quad (20)$$

where

$$\mathbf{J} = [\mathbf{X}', \mathbf{Y}']', \quad \mathbf{M} = [\Xi', \mathbf{Z}']', \quad \mathbf{L}_M = \begin{bmatrix} \Lambda_X & 0 \\ \Gamma(\mathbf{I} - \mathbf{B})^{-1}\Lambda_Y & (\mathbf{I} - \mathbf{B})^{-1}\Lambda_Y \end{bmatrix} \quad \text{and} \quad \mathbf{T} = [\mathbf{F}' \ \mathbf{E}']'.$$

“Since no necessary and sufficient conditions for identification are available”, Jöreskog (1981) suggests that “the identification problem be studied on a case by case basis, examining the equations, choosing the restrictions, not only in number but also in position”, in order to obtain unique parameters. Moreover, it has been demonstrated that even if the parameters of the LISREL model are perfectly identified, the scores of the LVs are not unique (Guttman, 1955; Schönemann, 1971; Vittadini, 1989). Given that the infinite scores of each LV in  $\mathbf{M}$  express the same concept, it would be opportune if these scores were highly and positively correlated. In practice, the correlations between the “two maximally different” solutions  $\mathbf{M}$  and  $\mathbf{M}^*$ , given by the diagonal elements of matrix

$$2\mathbf{D}_M^{-1/2} \Sigma_M \mathbf{L}_M \Sigma_J^{-1} \mathbf{L}'_M \Sigma_M \mathbf{D}_M^{-1/2} - \mathbf{D}_M^{-1/2} \Sigma_M \mathbf{D}_M^{-1/2} \quad (21)$$

(where  $\mathbf{D}_M$  is the diagonal matrix of  $\Sigma_M$ ) can be weak or even negative.

#### 4. The redundancy analysis approach to path modeling

In order to obtain solutions which respect properties (i)–(v), we propose a new approach based on the optimization of the redundancy criterion (Stewart and Love, 1968) which is entitled RA-PM.

Traditional redundancy analysis (Van den Wollenberg, 1977) does not calculate errors in equations and endogenous LVs, because it does not refer to structural equation models. RA-PM is an extension of redundancy analysis to a formative-reflective structural equation model with  $r$  formative blocks and  $s$  reflective blocks. It consists of an iterative procedure which is characterized by the following steps:

1. *Estimation of exogenous LVs and initial estimation of endogenous LVs*: The initialization of the RA-PM algorithm is given by estimating each exogenous LV as the first redundancy component of the  $\mathbf{Y}$  variables on the corresponding  $\mathbf{X}$  block. The choice of the redundancy criterion ensures that the maximum amount of information concerning  $\mathbf{Y}$  is extracted from each  $\mathbf{X}$  block and that the causal structure of the model is respected. Next, the initial estimates of the endogenous LVs are obtained by means of a Redundancy Analysis of each single  $\mathbf{Y}$  block on the estimates of the exogenous LVs.
2. *Estimation of further components and updated estimation of endogenous LVs*: The variables within each  $\mathbf{X}$  block are projected onto the orthogonal complement of the linear space spanned by all the redundancy components extracted previously. Thus, an orthogonal subspace is identified for each  $\mathbf{X}$  block. These subspaces are referred to as “residual  $\mathbf{X}$  blocks”. A second set of components is estimated by means of a redundancy analysis of the  $\mathbf{Y}$  variables on each residual  $\mathbf{X}$  block. Next, the endogenous LVs are re-estimated by means of a redundancy analysis of each single  $\mathbf{Y}$  block on the linear space spanned by the first and the second set of extracted components. This scheme is iterated according to a forward selection procedure. This procedure considers new residual blocks, extracts the corresponding redundancy components and updates the estimates of the endogenous LVs, until the accounted redundancy added at each iteration becomes marginal.

3. *Refinement of the solutions:* A backward selection procedure of the extracted components is performed in order to identify the minimal subset of components which accounts for a given  $\mathbf{Y}$  redundancy level. Note that although backward selection requires some amount of computational work, it guarantees that we obtain a more parsimonious model. On the basis of the components which have been selected by backward selection, the final estimates of the endogenous LVs are computed.
4. *Estimation of structural parameters and loadings:* Finally, the structural parameters of the model are estimated, together with the errors in equations. Moreover, the estimation of the loadings of each  $\mathbf{y}$  variable on the corresponding endogenous LV is provided.

In the following, we present each step analytically, assuming that the MVs are centred.

4.1. *Estimation of exogenous LVs and initial estimation of endogenous LVs*

*Estimation of exogenous LVs:* Each exogenous LV  $\xi_{(j)}$ ,  $j = 1, \dots, r$ , is estimated as the first redundancy component

$$\hat{\xi}_{(j)} = \mathbf{X}_{(j)} \hat{\omega}_{(j)}, \tag{22}$$

i.e. the unit variance linear combination of block  $\mathbf{X}_{(j)}$  which maximizes the redundancy index  $R_{(j)}^{(\xi)}$ , defined as

$$R_{(j)}^{(\xi)} = \sum_{k=1}^q \text{cor}(\mathbf{X}_{(j)} \omega_{(j)}, \mathbf{y}_k)^2 / q. \tag{23}$$

In this way,  $\hat{\omega}_{(j)}$  is computed as the eigenvector corresponding to the highest eigenvalue of matrix  $\Sigma_{\mathbf{X}_{(j)} \mathbf{X}_{(j)}}^{-1} \Sigma_{\mathbf{X}_{(j)} \mathbf{Y}} \Sigma_{\mathbf{Y} \mathbf{X}_{(j)}}$ , under the constraint

$$\hat{\omega}_{(j)}' \Sigma_{\mathbf{X}_{(j)} \mathbf{X}_{(j)}} \hat{\omega}_{(j)} = 1 \tag{24}$$

in such a way that  $\hat{\xi}_{(j)}$  is standardized. We have thus established the following association:

$$\begin{array}{lcl} \mathbf{X}_{(1)} & \rightarrow & \hat{\xi}_{(1)} \\ \vdots & & \vdots \\ \mathbf{X}_{(r)} & \rightarrow & \hat{\xi}_{(r)}. \end{array}$$

The estimates of the exogenous LVs are arranged in matrix  $\hat{\Xi}$ , defined as

$$\hat{\Xi} = [\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}]. \tag{25}$$

*Initial estimation of endogenous LVs:* For each block  $\mathbf{Y}_{(k)}$ ,  $k = 1, \dots, s$ , each endogenous LV  $\eta_{(k)}$  is estimated as the first redundancy component  $\hat{\eta}_{(k)}^{(\text{in})}$  on the linear space  $V(\hat{\Xi})$  spanned by the columns of  $\hat{\Xi}$ , i.e. the unit variance linear combination of the estimated exogenous LVs which maximizes the redundancy index  $R_{(k)}^{(\eta^{(\text{in})})}$ , defined as

$$R_{(k)}^{(\eta^{(\text{in})})} = \sum_{i=1}^{q_k} \text{cor}(\hat{\Xi} \rho_{(k)}, \mathbf{y}_{(k)i})^2 / q_k. \tag{26}$$

In this way,  $\hat{\rho}_{(k)}$  is the eigenvector corresponding to the highest eigenvalue of matrix  $\Sigma_{\hat{\Xi} \hat{\Xi}}^{-1} \Sigma_{\hat{\Xi} \mathbf{Y}_{(k)}} \Sigma_{\mathbf{Y}_{(k)} \hat{\Xi}}$ , under the constraint

$$\hat{\rho}_{(k)}' \Sigma_{\hat{\Xi} \hat{\Xi}} \hat{\rho}_{(k)} = 1 \tag{27}$$

in such a way that  $\hat{\eta}_{(k)}$  is standardized.

We have thus established the following association:

$$\begin{array}{lcl} \mathbf{Y}_{(1)} & \rightarrow & \hat{\eta}_{(1)}^{(\text{in})} \\ \vdots & & \vdots \\ \mathbf{Y}_{(s)} & \rightarrow & \hat{\eta}_{(s)}^{(\text{in})}. \end{array}$$

4.2. Estimation of further components and updated estimation of endogenous LVs

*Estimation of further components:* We define  $V^{(X)}$  as the linear space spanned by  $\mathbf{X}_{(1)}, \dots, \mathbf{X}_{(r)}$ ,  $V^{(1)}$  as the linear space spanned by  $\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}$ ,  $\mathbf{P}_{V^{(1)}}$  and  $\mathbf{Q}_{V^{(1)}} = \mathbf{I} - \mathbf{P}_{V^{(1)}}$  as the orthogonal projectors onto  $V^{(1)}$  and  $V^{(1)\perp}$ , respectively (here the orthogonal complement is defined in  $V^{(X)}$ ). For  $j = 1, \dots, r$ , the “residual blocks”  $\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}$  are taken into consideration. The second component of block  $\mathbf{X}_{(j)}$  is estimated as the first redundancy component, i.e. the unit variance linear combination

$$\hat{\phi}_{(j)}^{(1)} = \mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}\psi_{(j)}, \quad j = 1, \dots, r \tag{28}$$

which maximizes the redundancy index  $R_{(j)}^{(\phi)}$ , defined as

$$R_{(j)}^{(\phi)} = \sum_{k=1}^q \text{cor}(\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}\psi_{(j)}, \mathbf{y}_k)^2 / q. \tag{29}$$

In this way,  $\hat{\psi}_{(j)}$  is the eigenvector corresponding to the highest eigenvalue of matrix

$$\Sigma_{\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}}^{-1} \Sigma_{\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}} \mathbf{Y} \Sigma_{\mathbf{Y}\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}}$$

under the constraint

$$\hat{\psi}_{(j)}' \Sigma_{\mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)}} \mathbf{Q}_{V^{(1)}}\mathbf{X}_{(j)} \hat{\psi}_{(j)} = 1 \tag{30}$$

in such a way that  $\hat{\phi}_{(j)}$  is standardized.

We have thus established the following association:

$$\begin{matrix} \mathbf{X}_{(1)} & \rightarrow & \hat{\xi}_{(1)} & \hat{\phi}_{(1)}^{(1)} \\ \vdots & & \vdots & \vdots \\ \mathbf{X}_{(r)} & \rightarrow & \hat{\xi}_{(r)} & \hat{\phi}_{(r)}^{(1)}. \end{matrix}$$

It can be observed that the estimates  $\hat{\phi}_{(1)}^{(1)}, \dots, \hat{\phi}_{(r)}^{(1)}$  of the second set of components are not correlated to any of the estimates  $\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}$  of the exogenous LVs (i.e. the first set of components), being linear combinations of vectors in  $V^{(1)\perp}$ . In this way, we ensure that the new components share no part of the already extracted redundancy.

Next, we arrange the estimates  $\hat{\phi}_{(1)}^{(1)}, \dots, \hat{\phi}_{(r)}^{(1)}$  in matrix  $\hat{\Phi}^{(1)}$ , defined as

$$\hat{\Phi}^{(1)} = [\hat{\phi}_{(1)}^{(1)}, \dots, \hat{\phi}_{(r)}^{(1)}] \tag{31}$$

and introduce matrix  $\hat{\Pi}^{(1)}$ , where the first  $r$  columns are those of matrix  $\hat{\Xi}$  while the remaining columns are those of matrix  $\hat{\Phi}^{(1)}$

$$\hat{\Pi}^{(1)} = [\hat{\Xi}, \hat{\Phi}^{(1)}]. \tag{32}$$

*Updated estimation of endogenous LVs:* For each block  $\mathbf{Y}_{(k)}$ ,  $k = 1, \dots, s$ , the estimate  $\hat{\eta}_{(k)}$  is updated by computing the first redundancy component  $\hat{\eta}_{(k)}^{(\text{upd})}$  on the linear space  $V^{(1)}$  spanned by the columns of  $\hat{\Pi}^{(1)}$ , i.e. the unit variance linear combination of the extracted components, which maximizes the redundancy index  $R_{(k)}^{(\eta^{\text{upd}})}$ , defined as

$$R_{(k)}^{(\eta^{\text{upd}})} = \sum_{i=1}^{q_k} \text{cor}(\hat{\Pi}^{(1)}\theta_{(k)}, \mathbf{y}_{(k)i})^2 / q_k. \tag{33}$$

In this way,  $\hat{\theta}_{(k)}$  is the eigenvector corresponding to the highest eigenvalue of matrix  $\Sigma_{\hat{\Pi}^{(1)}\hat{\Pi}^{(1)}}^{-1} \Sigma_{\hat{\Pi}^{(1)}\mathbf{Y}_{(k)}} \Sigma_{\mathbf{Y}_{(k)}\hat{\Pi}^{(1)}}$ , under the constraint

$$\hat{\theta}_{(k)}' \Sigma_{\hat{\Pi}^{(1)}\hat{\Pi}^{(1)}} \hat{\theta}_{(k)} = 1 \tag{34}$$

in such a way that  $\hat{\eta}_{(k)}$  is standardized.

We have thus established the association:

$$\begin{matrix} \mathbf{Y}_{(1)} & \rightarrow & \hat{\boldsymbol{\eta}}_{(1)}^{(\text{upd})} \\ \vdots & & \vdots \\ \mathbf{Y}_{(s)} & \rightarrow & \hat{\boldsymbol{\eta}}_{(s)}^{(\text{upd})}. \end{matrix}$$

The above scheme can be iterated in order to provide the estimation of more sets of components  $\hat{\boldsymbol{\Phi}}^{(2)}, \dots, \hat{\boldsymbol{\Phi}}^{(p)}$  and consequently, to update the estimates  $\hat{\boldsymbol{\eta}}_{(1)}^{(\text{upd})}, \dots, \hat{\boldsymbol{\eta}}_{(s)}^{(\text{upd})}$  of the endogenous LVs.

In order to establish the number of components to be extracted and a suitable stopping rule, we adopt a forward selection procedure. This means that the newly extracted components will be utilized for updating the estimates of the endogenous LVs in the following two situations: first, if these new components add a relevant marginal proportion of  $\mathbf{Y}$  redundancy, or second, if the  $\mathbf{Y}$  block redundancy accounted for by the updated  $\hat{\boldsymbol{\eta}}^{(\text{upd})}$  variables increases in a significant way (see Section 5 for a discussion of the indicators to be used to perform these evaluations).

### 4.3. Refinement of the solutions

*Backward selection of the extracted components:* In order to achieve the most parsimonious model, we apply a backward selection procedure on the set of the extracted components: for each block  $\mathbf{X}_{(j)}$  we eliminate the components  $\hat{\boldsymbol{\phi}}_{(j)}$  whose removal has minor effect on the accounted  $\mathbf{Y}$  redundancies. This makes it possible to obtain a different number of components for each block  $\mathbf{X}_{(j)}$

$$\begin{matrix} \mathbf{X}_{(1)} & \rightarrow & \hat{\boldsymbol{\xi}}_{(1)} & \hat{\boldsymbol{\phi}}_{(1)}^{(1)} & \dots & \hat{\boldsymbol{\phi}}_{(1)}^{(v_1)} \\ \vdots & & \vdots & \vdots & \vdots & \\ \mathbf{X}_{(r)} & \rightarrow & \hat{\boldsymbol{\xi}}_{(r)} & \hat{\boldsymbol{\phi}}_{(r)}^{(1)} & \dots & \dots & \hat{\boldsymbol{\phi}}_{(r)}^{(v_r)}, \end{matrix}$$

where  $v_j, j = 1, \dots, r$ , is the number of selected components for block  $\mathbf{X}_{(j)}$ .

The estimates of the selected components are arranged in matrix  $\hat{\boldsymbol{\Phi}}$

$$\hat{\boldsymbol{\Phi}} = [\hat{\boldsymbol{\phi}}_{(1)}^{(1)}, \dots, \hat{\boldsymbol{\phi}}_{(r)}^{(1)}, \dots, \hat{\boldsymbol{\phi}}_{(1)}^{(v_1)}, \dots, \hat{\boldsymbol{\phi}}_{(r)}^{(v_r)}]. \tag{35}$$

We then introduce matrix  $\hat{\boldsymbol{\Pi}}$  whose first  $r$  columns are those of matrix  $\hat{\boldsymbol{\Xi}}$  while the remaining columns are those of matrix  $\hat{\boldsymbol{\Phi}}$

$$\hat{\boldsymbol{\Pi}} = [\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Phi}}], \tag{36}$$

thus the columns of  $\hat{\boldsymbol{\Pi}}$  contain all the components which have been selected by means of backward selection.

Finally, we indicate by  $V^{(\boldsymbol{\Pi})}$  the linear space spanned by the columns of  $\hat{\boldsymbol{\Pi}}$ .

*Final estimation of endogenous LVs:* For each block  $\mathbf{Y}_{(k)}, k = 1, \dots, s$ , the final estimate  $\hat{\boldsymbol{\eta}}_{(k)}$  is obtained as the first redundancy component on the linear subspace  $V^{(\boldsymbol{\Pi})}$ , i.e. the unit variance linear combination of the selected components which maximizes the redundancy index  $R_{(k)}^{(\eta)}$ , defined as

$$R_{(k)}^{(\eta)} = \sum_{i=1}^{q_k} \text{cor}(\hat{\boldsymbol{\Pi}}\boldsymbol{\theta}_{(k)}, \mathbf{y}_{(k)i})^2 / q_k. \tag{37}$$

In this way,  $\hat{\boldsymbol{\theta}}_{(k)}$  is the eigenvector corresponding to the highest eigenvalue of matrix  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\Pi}}^{-1}}^{-1} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\Pi}}\mathbf{Y}_{(k)}} \boldsymbol{\Sigma}_{\mathbf{Y}_{(k)}\hat{\boldsymbol{\Pi}}}$ , under the constraint

$$\hat{\boldsymbol{\theta}}_{(k)}' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\Pi}}} \hat{\boldsymbol{\theta}}_{(k)} = 1 \tag{38}$$

in such a way that  $\hat{\boldsymbol{\eta}}_{(k)}$  is standardized.

We have thus established the final association:

$$\begin{array}{ccc} \mathbf{Y}_{(1)} & \rightarrow & \hat{\boldsymbol{\eta}}_{(1)} \\ \vdots & & \vdots \\ \mathbf{Y}_{(s)} & \rightarrow & \hat{\boldsymbol{\eta}}_{(s)}. \end{array}$$

If we arrange the final estimates  $\hat{\boldsymbol{\eta}}_{(1)}, \dots, \hat{\boldsymbol{\eta}}_{(s)}$  in matrix  $\hat{\mathbf{H}} = [\hat{\boldsymbol{\eta}}_{(1)}, \dots, \hat{\boldsymbol{\eta}}_{(s)}]$  and the corresponding eigenvectors  $\hat{\boldsymbol{\theta}}_{(1)}, \dots, \hat{\boldsymbol{\theta}}_{(s)}$  in matrix  $\hat{\boldsymbol{\Theta}} = [\hat{\boldsymbol{\theta}}_{(1)}, \dots, \hat{\boldsymbol{\theta}}_{(s)}]$ , we can express the final estimates of the endogenous LVs as

$$\hat{\mathbf{H}} = \hat{\Pi} \hat{\boldsymbol{\Theta}} = \hat{\Xi} \hat{\boldsymbol{\Theta}}_1 + \hat{\Phi} \hat{\boldsymbol{\Theta}}_2, \quad (39)$$

where  $\hat{\boldsymbol{\Theta}}_1$  is the matrix obtained from  $\hat{\boldsymbol{\Theta}}$  retaining only the first  $r$  rows and  $\hat{\boldsymbol{\Theta}}_2$  the matrix obtained from  $\hat{\boldsymbol{\Theta}}$  retaining only the rows after the  $r$ th one.

#### 4.4. Structural parameters and loadings

*Structural parameters and errors in equations:* The estimate  $\hat{\gamma}_{(1)}$  of the structural parameter  $\gamma_{(1)}$  and the estimate  $\hat{\zeta}_{(1)}$  of the error in equation  $\zeta_{(1)}$  are obtained by means of a regression of  $\hat{\boldsymbol{\eta}}_{(1)}$  on  $\hat{\Xi}$

$$\hat{\boldsymbol{\eta}}_{(1)} = \hat{\Xi} \hat{\gamma}_{(1)} + \hat{\zeta}_{(1)}, \quad (40)$$

where  $\hat{\gamma}_{(1)}$  is the first column of matrix  $\hat{\boldsymbol{\Theta}}_1$  and  $\hat{\zeta}_{(1)}$  can be decomposed as  $\hat{\zeta}_{(1)} = \hat{\Phi} \hat{\alpha}_{(1)}$ , with  $\hat{\alpha}_{(1)}$  as the first column of  $\hat{\boldsymbol{\Theta}}_2$ .

Under the assumption that  $\hat{\boldsymbol{\eta}}_{(1)}$  and the columns  $\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}$  of  $\hat{\Xi}$  form a linearly independent set of regressors, the estimates  $\hat{\gamma}_{(2)}$  and  $\hat{\zeta}_{(2)}$  are obtained by means of a regression of  $\hat{\boldsymbol{\eta}}_{(2)}$  on  $\hat{\boldsymbol{\eta}}_{(1)}$  and  $\hat{\Xi}$ :

$$\hat{\boldsymbol{\eta}}_{(2)} = \hat{\boldsymbol{\eta}}_{(1)} \hat{b}_{(2)} + \hat{\Xi} \hat{\gamma}_{(2)} + \hat{\zeta}_{(2)}, \quad (41)$$

where  $\hat{\zeta}_{(2)}$  can be decomposed as  $\hat{\zeta}_{(2)} = \hat{\Phi} \hat{\alpha}_{(2)}$ .

For  $k = 3, \dots, s$ , under the assumption that the columns  $\hat{\boldsymbol{\eta}}_{(1)}, \dots, \hat{\boldsymbol{\eta}}_{(k-1)}$  of matrix  $\hat{\mathbf{H}}_{(k-1)}$  and the columns of  $\hat{\Xi}$  form a linearly independent set of regressors, the estimates  $\hat{\gamma}_{(k)}$  and  $\hat{\zeta}_{(k)}$  are obtained by means of a regression of  $\hat{\boldsymbol{\eta}}_{(k)}$  on  $\hat{\mathbf{H}}_{(k-1)}$  and  $\hat{\Xi}$

$$\hat{\boldsymbol{\eta}}_{(k)} = \hat{\mathbf{H}}_{(k-1)} \hat{\mathbf{b}}_{(k)} + \hat{\Xi} \hat{\gamma}_{(k)} + \hat{\zeta}_{(k)}, \quad k = 3, \dots, s, \quad (42)$$

where  $\hat{\zeta}_{(k)}$  can be decomposed as  $\hat{\zeta}_{(k)} = \hat{\Phi} \hat{\alpha}_{(k)}$ .

We now define  $\hat{\boldsymbol{\beta}}_{(1)}$  as a zero vector and  $\hat{\boldsymbol{\beta}}_{(k)}$ ,  $k = 2, \dots, s$ , by means of the following association:

$$\hat{\beta}_{(k)i} = \hat{b}_{(k)i}, \quad 1 \leq i \leq k-1 \quad \text{or} \quad \hat{\beta}_{(k)i} = 0, \quad k \leq i \leq s, \quad (43)$$

where  $\hat{b}_{(k)i}$  is the  $i$ th component of vector  $\hat{\mathbf{b}}_{(k)}$ .

If we consider vectors  $\hat{\boldsymbol{\beta}}_{(k)}$ ,  $\hat{\boldsymbol{\gamma}}_{(k)}$  and  $\hat{\zeta}_{(k)}$ ,  $k = 1, \dots, s$ , as the  $k$ th columns of matrices  $\hat{\mathbf{B}}$ ,  $\hat{\boldsymbol{\Gamma}}$  and  $\hat{\mathbf{Z}}$ , respectively, we can express the final estimates of the endogenous LVs in the form of Eq. (9)

$$\hat{\mathbf{H}} = \hat{\mathbf{H}} \hat{\mathbf{B}} + \hat{\Xi} \hat{\boldsymbol{\Gamma}} + \hat{\mathbf{Z}}. \quad (44)$$

*Loadings:* For each variable  $\mathbf{y}_{(k)i}$ ,  $k = 1, \dots, s$ ,  $i = 1, \dots, q_k$ , the loadings on the corresponding variable  $\hat{\boldsymbol{\eta}}_{(k)}$  are estimated as

$$\hat{\lambda}_{(k)i} = (\hat{\boldsymbol{\eta}}'_{(k)} \hat{\boldsymbol{\eta}}_{(k)})^{-1} \hat{\boldsymbol{\eta}}'_{(k)} \mathbf{y}_{(k)i}. \quad (45)$$

### 5. Model validation

At this point, an important question is raised: does the introduction of the errors in equations improve the model? And if so, to what degree?

In order to evaluate the effectiveness of RA-PM as compared to traditional PLS-PM, we utilize the redundancy criterion for both the exogenous and endogenous LVs.

An absolute measure of goodness of fit for the estimates of the exogenous LVs is the accounted redundancy associated to the linear subspace  $V(\Xi)$  with respect to  $\mathbf{Y}$

$$\text{Red}(\mathbf{Y}|V(\Xi)) = \frac{\text{Tr}(\Sigma_{\hat{\Xi}\hat{\Xi}}^{-1} \Sigma_{\hat{\Xi}\mathbf{Y}} \Sigma_{\mathbf{Y}\hat{\Xi}})}{\text{Tr}(\mathbf{Y}'\mathbf{Y})}. \tag{46}$$

By dividing  $\text{Red}(\mathbf{Y}|V(\Xi))$  by its theoretical maximum, we obtain a relative measure of goodness of fit

$$\text{Red}^{\text{REL}}(\mathbf{Y}|V(\Xi)) = \frac{\text{Tr}(\Sigma_{\hat{\Xi}\hat{\Xi}}^{-1} \Sigma_{\hat{\Xi}\mathbf{Y}} \Sigma_{\mathbf{Y}\hat{\Xi}})}{\text{Tr}(\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{X}})}. \tag{47}$$

If

$$\text{Red}(\mathbf{Y}|\mathbf{X}) = \frac{\text{Tr}(\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{X}})}{\text{Tr}(\mathbf{Y}'\mathbf{Y})} \tag{48}$$

is the redundancy accounted for by  $\mathbf{X}$ , from (46) and (47) we have

$$\text{Red}(\mathbf{Y}|V(\Xi)) = \text{Red}^{\text{REL}}(\mathbf{Y}|V(\Xi)) \cdot \text{Red}(\mathbf{Y}|\mathbf{X}). \tag{49}$$

Note that an analogous measure of goodness of fit may be calculated for the extracted components, thus representing the accounted redundancy associated to the linear subspace  $V(\Pi)$  with respect to  $\mathbf{Y}$ .

A measure of goodness of fit for the estimates of the endogenous LVs is the redundancy accounted for by  $\hat{\boldsymbol{\eta}}_{(k)}$  with respect to the single block  $\mathbf{Y}_{(k)}$

$$\text{Red}(\mathbf{Y}_{(k)}|\hat{\boldsymbol{\eta}}_{(k)}) = \frac{\hat{\boldsymbol{\eta}}'_{(k)} \mathbf{Y}_{(k)} \mathbf{Y}'_{(k)} \hat{\boldsymbol{\eta}}_{(k)}}{\text{Tr}(\mathbf{Y}'_{(k)} \mathbf{Y}_{(k)})}, \quad k = 1, \dots, s. \tag{50}$$

### 6. RA-PM properties

#### 6.1. The utilization of residual information in the formative model

Tenenhaus et al. (2005) state that “in a formative model, the block of MVs can be multidimensional”. In practice, multidimensionality can be verified by means of various indicators (e.g. Goldstein-Dillon’s  $\rho$ , Cronbach’s  $\alpha$  or Principal Component Analysis). In these circumstances, in order to extract the residual information in the data, more redundancy components need to be considered. In RA-PM, this is achieved by an iterative algorithm that extracts the most informative components and generates a suitable orthogonal decomposition of the linear spaces spanned by the MVs in each  $\mathbf{X}$  block. In practice, on the basis of the redundancy criterion, the algorithm starts by extracting the most informative component within each  $\mathbf{X}$  block. By means of suitable orthogonal projections, it then identifies the “residual subspaces” (one for each  $\mathbf{X}$  block), i.e. those containing information which has not yet been extracted. At each successive iteration, the most informative component within each residual subspace is extracted and new residual subspaces are obtained. The algorithm stops when all the relevant residual information about  $\mathbf{Y}$  contained in the  $\mathbf{X}$  blocks is extracted.

#### 6.2. Algebraic description and optimality

In order to justify the backward selection procedure and demonstrate the optimality of RA-PM algorithm, we adopt an algebraic perspective based on the orthogonal decomposition of the linear space  $V(\mathbf{X})$  induced by the iterative extraction of the redundancy components.

By performing the estimation of the exogenous LVs, we identify the linear subspace  $V^{(\Xi)} \subset V^{(\mathbf{X})}$ , defined as the linear span of  $\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}$ . Similarly, when a second set of components is extracted, the linear subspace  $V^{(1)} \subset V^{(\mathbf{X})}$  is identified as the linear span of  $\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}, \hat{\phi}_{(1)}^{(1)}, \dots, \hat{\phi}_{(r)}^{(1)}$ .

The components which are extracted at different iterations are uncorrelated, thus the linear subspace  $W^{(1)} = \text{span}(\hat{\phi}_{(1)}^{(1)}, \dots, \hat{\phi}_{(r)}^{(1)})$  is orthogonal to  $V^{(\Xi)}$  and the linear subspace  $V^{(1)}$  can be expressed as

$$V^{(1)} = V^{(\Xi)} \oplus W^{(1)}. \quad (51)$$

As new sets of redundancy components are extracted, a chain of subspaces  $V^{(\Xi)} \subseteq V^{(1)} \subseteq \dots \subseteq V^{(\ell)} \subseteq V^{(\mathbf{X})}$ ,  $\ell = \max(v_1, \dots, v_r)$ , is generated.

From the above discussion it can be deduced that

$$V^{(k)} = V^{(k-1)} \oplus W^{(k)}, \quad k = 1, \dots, \ell, \quad (52)$$

where  $W^{(k)}$  is the linear space spanned by the redundancy components extracted during iteration  $k$ , and that  $V^{(\mathbf{X})}$  can be decomposed as

$$V^{(\mathbf{X})} = V^{(\Xi)} \oplus W^{(1)} \oplus \dots \oplus W^{(\ell)}. \quad (53)$$

At this point, if we halt the extractions at iteration  $\hat{\ell} \leq \ell$  according to forward selection (i.e. when the marginal contribution to the accounted redundancies due to  $W^{(\hat{\ell}+1)}$  is negligible),  $V^{(\mathbf{X})}$  can be decomposed as

$$V^{(\mathbf{X})} = V^{(\hat{\ell})} \oplus V^{(\hat{\ell})^\perp}. \quad (54)$$

If we then establish a target level of  $\mathbf{Y}$  redundancy to be accounted for by the extracted components, backward selection provides the removal of the less contributing components, reducing  $V^{(\hat{\ell})}$  to a subspace  $V^{(\Pi)}$  of minimum dimension accounting for the target redundancy level. In this sense, the procedure of estimation and selection of the components (i.e. the procedure of identification of the subspace  $V^{(\Pi)}$ ) is optimal. Based on this selection, the computation of the final estimates of the endogenous LVs fulfills the optimality criterion of redundancy analysis. Hence, we can affirm that the proposed estimation procedure for the LVs is optimal.

### 6.3. Features of RA-PM estimates

As a consequence of the adoption of the RA-PM algorithm, the following algebraic and geometrical features hold:

- (i) Once the exogenous and endogenous LVs have been estimated, the estimates  $\hat{\mathbf{B}}$  and  $\hat{\Gamma}$  of the structural parameters and the estimates  $\hat{\Lambda}_Y$  of the loadings are univocally determined.
- (ii) The estimates  $\hat{\mathbf{Z}}$  of the errors in equations are linear combinations of the estimates  $\hat{\Phi}$  of the further components, which are uncorrelated with the estimates  $\hat{\Xi}$  of the exogenous LVs; it follows that

$$\text{cov}(\hat{\xi}_{(j)}, \hat{\zeta}_{(k)}) = 0, \quad j = 1, \dots, r, \quad k = 1, \dots, s, \quad (55)$$

meaning that the estimates  $\hat{\mathbf{Z}}$  share no information with the estimates  $\hat{\Xi}$ .

- (iii) The estimates  $\hat{\Xi}$  of the exogenous LVs are univocally defined as linear combinations of the  $\mathbf{X}$  variables, the estimates  $\hat{\mathbf{H}}$  of the endogenous LVs and the estimates  $\hat{\mathbf{Z}}$  of the errors in equations are univocally defined as linear combinations of the estimates  $\hat{\Xi}$  and  $\hat{\Phi}$ . Moreover, the rank of the set  $\{\hat{\xi}_{(1)}, \dots, \hat{\xi}_{(r)}, \hat{\zeta}_{(1)}, \dots, \hat{\zeta}_{(s)}\}$  does not exceed the rank of  $\mathbf{X}$ .
- (iv) The solutions respect the causal structure of the model. In fact, in the RA-PM algorithm the estimates  $\hat{\Xi}$  are linear combinations of the  $\mathbf{X}$  variables, the estimates  $\hat{\mathbf{Z}}$  are linear combinations of the estimates  $\hat{\Phi}$ , and the estimates  $\hat{\mathbf{H}}$  are linear combinations of the estimates  $\hat{\Xi}$  and  $\hat{\mathbf{Z}}$ .
- (v) The errors in equations are taken into consideration for the estimation of the endogenous LVs, improving their redundancy accounting capability.

Table 1  
LVs and MVs in the HC model

Exogenous LVs	Formative MVs
Educational human capital ( $\xi_1$ )	Educational indicators H Years of schooling ( $z_1$ ) S Years of schooling ( $z_2$ ) H Father's years of schooling ( $z_3$ ) H Mother's years of schooling ( $z_4$ ) S Father's years of schooling ( $z_5$ ) S Mother's years of schooling ( $z_6$ ) Household real assets ( $z_7$ ) Household financial assets ( $z_8$ ) Household total debt ( $z_9$ )
Work experience human capital ( $\xi_2$ )	Job indicators H Age of entrance in the labour market ( $x_1$ ) S Age of entrance in the labour market ( $x_2$ ) H Number of years of full time job ( $x_3$ ) S Number of years of full time job ( $x_4$ ) H Number of years of part time job ( $x_5$ ) S Number of years of part time job ( $x_6$ )
Endogenous LVs	Reflective MVs
Ability to generate earned income ( $\eta_1$ )	Earned income Household net disposable earned income (wages and salaries) ( $y_1$ ) Household pensions and net transfers ( $y_2$ ) Household net income from self-employment ( $y_3$ )
Ability to generate property income ( $\eta_2$ )	Property income Household income from property ( $y_4$ ) Household financial income ( $y_5$ )

Table 2  
Percentage of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  redundancy accounted for by  $\mathbf{X}$

Red( $\mathbf{Y} \mathbf{X}$ )	Red( $\mathbf{Y}_{(1)} \mathbf{X}$ )	Red( $\mathbf{Y}_{(2)} \mathbf{X}$ )
34.51	31.86	60.68

## 7. The HC model

### 7.1. The variables

We will now briefly describe the MVs involved in the HC model introduced in Section 1. The data regard  $n = 1936$  Italian families and are drawn from the survey on Italian Household Budget performed by Banca d'Italia (2002).

As indicated in Table 1, the formative MVs associated with the two exogenous LVs regard variables pertaining to education and work experience. Similarly, the reflective MVs associated with the two endogenous LVs regard the indicators of earned and property income. For each family in the survey, some of the above variables are measured for both household head (H) and spouse (S).

### 7.2. The model estimation

Before estimating the exogenous and endogenous LVs involved in the HC model, we evaluated the redundancy accounting capability of  $\mathbf{X}$  with respect to  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$ . The results are reported in Table 2.

We then analysed the redundancy structure of the data. It is known that each eigenvalue of matrix  $\Sigma_{\mathbf{P}_{\mathbf{X}_{(j)}} \mathbf{Y} \mathbf{P}_{\mathbf{X}_{(j)}}}$  is associated with a redundancy component of  $\mathbf{X}_{(j)}$  with respect to  $\mathbf{Y}$  (D'Ambra and Lauro, 1982); the analysis of these eigenvalues enables to formulate some hypotheses on the optimal number of redundancy components to be extracted (see Table 3).

Table 3  
First five eigenvalues for  $\mathbf{X}_{(1)}$  and  $\mathbf{X}_{(2)}$  blocks

	1st	2nd	3rd	4th	5th
$\mathbf{X}_{(1)}$ block	0.9010	0.0901	0.0071	0.0013	0.0005
$\mathbf{X}_{(2)}$ block	0.5161	0.3366	0.1208	0.0233	0.0031

Table 4  
Percentage of  $\mathbf{Y}$  redundancy accounted for by the extracted components at different iterations

Iteration	Red( $\mathbf{Y} V^{(\Psi)}$ )	Red <sup>REL</sup> ( $\mathbf{Y} V^{(\Psi)}$ )
1	21.02	61.49
2	30.92	89.61
3	34.07	98.72
4	34.40	99.67
5	34.48	99.91

Table 5  
Percentage of  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  accounted redundancy for by  $\hat{\boldsymbol{\eta}}_{(1)}$  and  $\hat{\boldsymbol{\eta}}_{(2)}$  at different iterations

Iteration	Estimate	Red( $\mathbf{Y}_{(1)} \hat{\boldsymbol{\eta}}_{(1)}$ )	Estimate	Red( $\mathbf{Y}_{(2)} \hat{\boldsymbol{\eta}}_{(2)}$ )
1	$\hat{\boldsymbol{\eta}}_{(1)}^{(in)}$	18.50	$\hat{\boldsymbol{\eta}}_{(2)}^{(in)}$	16.28
2	$\hat{\boldsymbol{\eta}}_{(1)}^{upd}$	18.96	$\hat{\boldsymbol{\eta}}_{(2)}^{upd}$	30.52
3	$\hat{\boldsymbol{\eta}}_{(1)}^{upd}$	19.30	$\hat{\boldsymbol{\eta}}_{(2)}^{upd}$	38.49

Table 6  
Percentage of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  redundancy accounted for by the RA-PM estimates

Red( $\mathbf{Y} V^{(\Pi)}$ )	Red( $\mathbf{Y}_{(1)} \hat{\boldsymbol{\eta}}_{(1)}$ )	Red( $\mathbf{Y}_{(2)} \hat{\boldsymbol{\eta}}_{(2)}$ )
32.43	18.79	38.03

It is evident that it is possible to extract only one redundancy component from block  $\mathbf{X}_{(1)}$  and three redundancy components from block  $\mathbf{X}_{(2)}$ . Note that if we do not use the components beyond the first extraction, we lose a relevant amount of information, as it is confirmed in Table 4, which reports the accounted  $\mathbf{Y}$  redundancy corresponding to different iterations of the algorithm. As the number of iterations grows, the accounted redundancy increases; however, the increasing rate slows down after five iterations. (Note that we have indicated by  $V^{(\Psi)}$  the linear space spanned by the extracted components).

At each iteration we computed the estimates of the endogenous LVs and the corresponding accounted  $\mathbf{Y}$  block redundancy (see Table 5). The relevance of the residuals (taken into account by means of the components extracted after the first iteration) in improving the accounted redundancy is highly evident, particularly for the estimate of  $\boldsymbol{\eta}_{(2)}$ .

By performing a backward selection procedure, and removing the less contributing variables, we ultimately selected  $\hat{\boldsymbol{\xi}}_{(1)}$ , the first component from  $\mathbf{X}_{(1)}$ , together with  $\hat{\boldsymbol{\xi}}_{(2)}$ ,  $\hat{\boldsymbol{\phi}}_{(2)}^{(1)}$  and  $\hat{\boldsymbol{\phi}}_{(2)}^{(2)}$ , (i.e. the first three components from  $\mathbf{X}_{(2)}$ ). Table 6 reports the redundancy accounted for by the selected components and the final estimates of the endogenous LVs. (We remind that  $V^{(\Pi)}$  indicates the linear space spanned by the selected components).

Table 7  
Percentage of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  redundancy accounted for by the PLS-PM estimates

$\text{Red}(\mathbf{Y}   \hat{\xi}_{(1)}^{\text{PLS}}, \hat{\xi}_{(2)}^{\text{PLS}})$	$\text{Red}(\mathbf{Y}_{(1)}   \hat{\eta}_{(1)}^{\text{PLS}})$	$\text{Red}(\mathbf{Y}_{(2)}   \hat{\eta}_{(2)}^{\text{PLS}})$
18.97	39.36	43.44

Table 8  
Jöreskog goodness of fit index (GFI) and minimum correlations

GFI	Minimum correlation			
	$\hat{\xi}_{(1)}^{\text{Lsr}}$	$\hat{\xi}_{(2)}^{\text{Lsr}}$	$\hat{\zeta}_{(1)}^{\text{Lsr}}$	$\hat{\zeta}_2^{\text{Lsr}}$
0.86	0.45	0.22	-0.01	-0.12

Table 9  
RA-PM outer weights

	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$
$\hat{\xi}_{(1)}$	0.41	0.51	-0.01	0.07	-0.04	0.13	0.29	0.01	-0.03
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$			
$\hat{\xi}_{(2)}$	-0.06	0.78	0.37	0.25	0.44	0.49			

### 7.3. Comparison with PLS path modeling and LISREL model

The HC data set has been also analysed by means of PLS-PM. The redundancy of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  which is accounted for by the PLS-PM estimates is reported in Table 7.

From the above, it is evident that, when only the first extraction is performed,  $\hat{\xi}_{(1)}^{\text{PLS}}$  and  $\hat{\xi}_{(2)}^{\text{PLS}}$  account for approximately the same  $\mathbf{Y}$  redundancy as RA-PM (21.02%, see Table 4). As more components are being estimated, the accounted  $\mathbf{Y}$  redundancy in RA-PM overcomes that accounted for by PLS-PM; this is particularly true for the selected components. It should be noted that the total  $\mathbf{Y}$  redundancy accounted for by  $\hat{\eta}_{(1)}^{\text{PLS}}$  and  $\hat{\eta}_{(2)}^{\text{PLS}}$  can be directly computed as 40.62%, while the total  $\mathbf{Y}$  redundancy accounted for by the  $\mathbf{X}$  variables is 34.51% (see Table 2). This demonstrates that a part of the accounting capability of PLS-PM is derived from the  $\mathbf{Y}$  variables themselves and not from the  $\mathbf{X}$  variables.

Before interpreting the results, we verified the consequences of the non-uniqueness of the scores of the LVs and the errors in equations in the LISREL Model. In Table 8, it can be seen that the goodness of fit of the LISREL Model is high, but the correlations between the two maximally different solutions regarding the same identified model for HC are far from 1 and also negative.

### 7.4. Interpretation of model parameters

The outer weights of the exogenous LV estimates (see Table 9) are in agreement with the knowledge of HC (Vittadini and Lovaglio, 2004); in fact,  $\hat{\xi}_{(1)}$  is correlated to both the years of schooling and household real assets (it thus represents the educational HC), and  $\hat{\xi}_{(2)}$  is correlated to years of work (it thus represents the work experience HC).

The structural parameters are reported in Table 10;  $\hat{\eta}_{(1)}$  expresses a contrast between educational HC and work experience HC, according to the Italian situation;  $\hat{\eta}_{(2)}$  presents structural parameters of different signs, but it is positively correlated with  $\hat{\xi}_{(1)}$ .

The loadings of each  $\mathbf{y}$  variable are reported in Table 11 and can be seen to be in agreement with a model for HC.

Table 10  
RA-PM structural parameters

	$\hat{\eta}_{(1)}$	$\hat{\eta}_{(2)}$
$\hat{\xi}_{(1)}$	0.34	1.58
$\hat{\xi}_{(2)}$	0.78	-1.92
$\hat{\eta}_{(1)}$	-	2.83

Table 11  
RA-PM loadings

	$y_{(1)1}$	$y_{(1)2}$	$y_{(1)3}$
$\hat{\eta}_{(1)}$	0.59	-0.47	0.11
	$y_{(2)1}$	$y_{(2)2}$	
$\hat{\eta}_{(2)}$	0.48	0.82	

### 8. RA-PM simulation study

#### 8.1. The simulation experiment

In order to illustrate the effectiveness of RA-PM, we have performed a simulation study based on model (7)–(9), with  $r = 2, s = 2, p = 8, q = 8$  and  $n = 1000$ . The purpose of the simulation is to show how taking into account redundancy components beyond the first extraction can improve the redundancy accounted for by the endogenous LV estimates, particularly when the covariance structure within and between the formative and the reflective MVs does not enable for unidimensional description of the variables. To this aim, we have first defined two reference situations, indicated by  $A$  and  $K$ .

In situation  $A$ , the variables in the  $\mathbf{Y}_{(1)}$  block are correlated with those in the  $\mathbf{X}_{(1)}$  block alone and, similarly, the variables in the  $\mathbf{Y}_{(2)}$  block are only correlated with those in the  $\mathbf{X}_{(2)}$  block; moreover the variables in the  $\mathbf{X}_{(1)}$  block are not correlated with the variables in the  $\mathbf{X}_{(2)}$  block.

On the contrary, in situation  $K$  there are cross correlations between the variables in the  $\mathbf{X}$  and the  $\mathbf{Y}$  blocks and there are correlations among some of the variables in the  $\mathbf{X}_{(1)}$  block and some of the variables in the  $\mathbf{X}_{(2)}$  block.

Situation  $A$  is transformed into situation  $K$  through nine intermediate situations (indicated by  $B$ – $J$ ). The transformation is driven by a parameter which progressively transforms  $\Sigma_{\mathbf{X}\mathbf{X}}^{(A)}$  and  $\Sigma_{\mathbf{X}\mathbf{Y}}^{(A)}$  into  $\Sigma_{\mathbf{X}\mathbf{X}}^{(K)}$  and  $\Sigma_{\mathbf{X}\mathbf{Y}}^{(K)}$ .

More specifically, we first defined two lower triangular matrices  $\mathbf{U}^{(A)}$  and  $\mathbf{U}^{(K)}$  such that matrices

$$\Sigma_{\mathbf{X}\mathbf{X}}^{(A)} = \mathbf{U}^{(A)}\mathbf{U}^{(A)'} \quad \text{and} \quad \Sigma_{\mathbf{X}\mathbf{X}}^{(K)} = \mathbf{U}^{(K)}\mathbf{U}^{(K)'} \tag{56}$$

reproduced the desired covariance structure for the two reference situations.

Starting from  $\mathbf{U}^{(A)}$  and  $\mathbf{U}^{(K)}$  we introduced a sequence of nine intermediate lower triangular matrices defined as

$$\mathbf{U}^{(t)} = (1 - t) \cdot \mathbf{U}^{(A)} + t \cdot \mathbf{U}^{(K)} \quad (t = 0.1, 0.2, \dots, 0.9). \tag{57}$$

After obtaining the sequence  $\{\mathbf{U}^{(t)}\}$ , the corresponding sequence  $\{\Sigma_{\mathbf{X}\mathbf{X}}^{(t)}\}$  was computed and nine data sets of  $\mathbf{X}$  variables, distributed as  $N(0, \Sigma_{\mathbf{X}\mathbf{X}}^{(t)})$ , were generated. In this paper we report the results for only the nine intermediate situations (from  $B$  to  $J$ ), given that the reference situations were used only as a starting point for the generation of the other data sets.

#### 8.2. Discussion of the results

For each simulation case, the eigenvalues of covariance matrices  $\Sigma_{\mathbf{P}_{\mathbf{X}_{(1)}}\mathbf{Y}_{\mathbf{P}_{\mathbf{X}_{(1)}}}\mathbf{Y}}$  and  $\Sigma_{\mathbf{P}_{\mathbf{X}_{(2)}}\mathbf{Y}_{\mathbf{P}_{\mathbf{X}_{(2)}}}\mathbf{Y}}$  indicate the number of relevant redundancy components within the  $\mathbf{X}$  variables (see Table 12). In both cases, it is evident that, moving from

Table 12  
First five eigenvalues/trace (percentage) for  $\mathbf{X}_{(1)}$  and  $\mathbf{X}_{(2)}$  blocks (situations  $B$ – $J$ )

	$\mathbf{X}_{(1)}$ block					$\mathbf{X}_{(2)}$ block				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
$B$	65.99	10.78	5.95	5.34	3.88	67.71	10.33	7.23	4.38	3.28
$C$	64.44	11.98	5.56	5.26	4.06	67.08	9.64	7.58	4.60	3.46
$D$	64.16	13.26	5.49	4.71	3.94	67.44	8.90	7.31	4.72	3.71
$E$	64.60	14.65	5.24	4.17	3.67	67.98	9.16	6.15	4.80	4.22
$F$	64.35	16.29	4.91	3.85	3.47	67.16	10.07	6.23	4.78	4.43
$G$	62.46	18.68	4.57	3.75	3.52	63.95	12.95	6.95	4.79	4.14
$H$	58.79	22.67	4.44	4.17	3.23	58.44	18.59	6.91	4.67	3.97
$I$	53.79	28.52	4.70	3.90	2.87	51.67	25.67	6.47	4.52	3.90
$J$	48.07	35.56	4.49	3.63	2.50	44.72	33.02	5.94	4.56	3.78

Table 13  
Percentage of accounted redundancy (situations  $B$ – $J$ )

	Red( $\mathbf{Y} \mathbf{X}$ )	Red( $\mathbf{Y} \mathbf{X}_{(1)}$ )	Red( $\mathbf{Y} \mathbf{X}_{(2)}$ )	Red( $\mathbf{Y}_{(1)} \mathbf{X}$ )	Red( $\mathbf{Y}_{(2)} \mathbf{X}$ )
$B$	64.58	32.67	32.69	64.50	64.66
$C$	61.43	32.16	31.72	61.84	61.02
$D$	58.90	32.63	31.45	59.91	57.89
$E$	57.27	34.03	31.94	58.79	55.75
$F$	56.80	36.27	33.33	58.51	55.08
$G$	57.54	39.13	35.77	59.04	56.04
$H$	59.34	42.34	39.24	60.34	58.34
$I$	61.86	45.63	43.41	62.34	61.37
$J$	64.77	48.91	47.89	64.96	64.57

situation  $B$  to situation  $J$ , a second component is progressively revealed. From an analytical point of view, this is due to the fact that the eigenvalues of a matrix are a continuous function of the matrix elements. In the simulation, the eigenvalues are a continuous function of parameter  $t$ , driving the computation of the covariance matrices from situation  $A$ , where there is only one relevant eigenvalue, to situation  $K$ , where there are two relevant eigenvalues.

Table 13 reports the results of a redundancy analysis performed on the variables involved in the simulation. It can be seen that the redundancy accounting capability of the  $\mathbf{X}$  variables concerning  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  decreases as we leave situation  $B$ , reaches a minimum in the middle cases, and then increases when approaching situation  $J$ ; on the contrary, the  $\mathbf{Y}$  redundancy accounting capability of  $\mathbf{X}_{(1)}$  and  $\mathbf{X}_{(2)}$  increases progressively from situation  $B$  to  $J$ .

We ran RA-PM on the nine data sets, extracting a pair of components (i.e. a component from each  $\mathbf{X}$  block) as the number of iterations grows.

Table 14 reports the  $\mathbf{Y}$  redundancy accounted for by the pairs of extracted components at each iteration. We note that the higher the proximity to situation  $J$  is reached, the less the first extraction (i.e. the estimation of the two exogenous LVs alone) accounts for the  $\mathbf{Y}$  redundancy. This justifies the extraction of more redundancy components.

In order to show the relevance of the components beyond the first extraction, Table 15 reports the  $\mathbf{Y}$  redundancy accounted for by both the initial estimates  $\hat{\eta}_{(1)}^{(in)}$  and  $\hat{\eta}_{(2)}^{(in)}$  and the final estimates  $\hat{\eta}_{(1)}$  and  $\hat{\eta}_{(2)}$ .

In the last three situations the improvement is remarkable, as it is expected by considering the redundancy structure of the simulation cases.

### 8.3. Comparison with PLS path modeling and LISREL model

The simulation data sets have been also analysed by means of PLS-PM. The redundancy of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  which is accounted for by the PLS-PM estimates in situations  $B$ – $J$  is reported in Table 16.

If in RA-PM only the first iteration is performed, the amount of  $\mathbf{Y}$  redundancy accounted for by the estimates of the two exogenous LVs is similar in the two approaches. However, if a second iteration is performed, the  $\mathbf{Y}$  redundancy

Table 14  
Percentage of  $\mathbf{Y}$  redundancy accounted for by the pairs of extracted components

	1	2	3	4	5	6	7	8
<i>B</i>	43.51	50.30	54.51	57.54	59.90	61.63	63.19	64.58
<i>C</i>	40.57	47.24	51.25	54.20	56.60	58.41	60.02	61.43
<i>D</i>	38.42	44.46	48.63	51.56	54.06	55.98	57.52	58.90
<i>E</i>	37.04	42.90	46.95	49.91	52.42	54.46	55.89	57.27
<i>F</i>	35.97	42.37	46.36	49.37	51.88	54.00	55.43	56.80
<i>G</i>	34.72	42.39	46.73	49.79	52.32	54.50	56.25	57.54
<i>H</i>	33.03	43.55	48.00	51.09	53.75	56.09	58.00	59.34
<i>I</i>	30.84	45.63	50.07	53.30	56.01	58.43	60.44	61.86
<i>J</i>	28.28	48.20	52.61	56.05	58.83	61.20	63.27	64.77

Table 15  
Percentage of  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  redundancy accounted for by  $\hat{\boldsymbol{\eta}}_{(1)}^{(in)}$ ,  $\hat{\boldsymbol{\eta}}_{(2)}^{(in)}$  and  $\hat{\boldsymbol{\eta}}_{(1)}$ ,  $\hat{\boldsymbol{\eta}}_{(2)}$

	$\text{Red}(\mathbf{Y}_{(1)} \hat{\boldsymbol{\eta}}_{(1)}^{(in)})$	$\text{Red}(\mathbf{Y}_{(1)} \hat{\boldsymbol{\eta}}_{(1)})$	$\text{Red}(\mathbf{Y}_{(2)} \hat{\boldsymbol{\eta}}_{(2)}^{(in)})$	$\text{Red}(\mathbf{Y}_{(2)} \hat{\boldsymbol{\eta}}_{(2)})$
<i>B</i>	42.93	42.96	43.74	43.74
<i>C</i>	39.78	39.85	40.23	40.27
<i>D</i>	37.12	37.24	37.08	37.20
<i>E</i>	35.11	35.63	34.47	34.76
<i>F</i>	33.56	35.11	32.34	33.19
<i>G</i>	32.09	35.77	30.46	33.17
<i>H</i>	30.45	37.61	28.54	35.09
<i>I</i>	28.72	40.18	26.23	38.22
<i>J</i>	27.73	43.30	22.70	41.87

Table 16  
Percentage of  $\mathbf{Y}$ ,  $\mathbf{Y}_{(1)}$  and  $\mathbf{Y}_{(2)}$  redundancy accounted for by the PLS-PM estimates

	$\text{Red}(\mathbf{Y} \hat{\boldsymbol{\xi}}_{(1)}^{\text{PLS}}, \hat{\boldsymbol{\xi}}_{(2)}^{\text{PLS}})$	$\text{Red}(\mathbf{Y}_{(1)} \hat{\boldsymbol{\eta}}_{(1)}^{\text{PLS}})$	$\text{Red}(\mathbf{Y}_{(2)} \hat{\boldsymbol{\eta}}_{(2)}^{\text{PLS}})$
<i>B</i>	47.78	51.02	52.26
<i>C</i>	44.90	48.41	48.56
<i>D</i>	42.35	46.11	43.34
<i>E</i>	40.09	44.50	36.86
<i>F</i>	37.85	43.83	29.64
<i>G</i>	35.38	44.18	22.47
<i>H</i>	32.74	45.44	16.12
<i>I</i>	30.23	47.34	11.15
<i>J</i>	28.18	49.49	7.65

accounted for by the RA-PM estimates becomes higher than in PLS-PM in all simulation cases. Concerning the estimates of the endogenous LVs, a direct comparison between RA-PM and PLS-PM is more problematic because these two approaches estimate the endogenous LVs in extremely different ways. In fact, while  $\hat{\boldsymbol{\eta}}_{(1)}^{\text{PLS}}$  and  $\hat{\boldsymbol{\eta}}_{(2)}^{\text{PLS}}$  are expressed as linear combinations of the  $\mathbf{Y}$  variables, the corresponding RA-PM estimates are expressed as linear combinations of the  $\mathbf{X}$  variables. Computing the redundancies of the  $\mathbf{Y}$  blocks in PLS-PM, we see that  $\hat{\boldsymbol{\eta}}_{(1)}^{\text{PLS}}$  accounts for a redundancy comparable with that accounted for by  $\hat{\boldsymbol{\eta}}_{(1)}$  in RA-PM. However,  $\hat{\boldsymbol{\eta}}_{(2)}^{\text{PLS}}$  reduces its accounting capability as it approaches situation *J*, where it accounts for 7.65%.

We conclude by noting that when the  $\mathbf{Y}$  redundancy structure of the formative blocks is not unidimensional, the addition of more redundancy components really improves the redundancy accounted capability of the RA-PM estimates of the endogenous LVs. We maintain that PLS-PM approach could also benefit from considering more

Table 17  
Jöreskog goodness of fit index (GFI) and minimum correlations

	GFI	Minimum correlation			
		$\hat{\xi}_{(1)}^{LSR}$	$\hat{\xi}_{(2)}^{LSR}$	$\hat{\xi}_{(1)}^{LSR}$	$\hat{\xi}_{(2)}^{LSR}$
<i>B</i>	0.85	0.68	0.65	0.68	0.68
<i>C</i>	0.84	0.62	0.63	0.64	0.64
<i>D</i>	0.82	0.54	0.60	0.62	0.60
<i>E</i>	0.80	0.50	0.56	0.60	0.57
<i>F</i>	0.78	0.42	0.52	0.59	0.55
<i>G</i>	0.76	0.38	0.47	0.59	0.54
<i>H</i>	0.72	0.34	0.41	0.60	0.54
<i>I</i>	0.68	0.31	0.35	0.61	0.57
<i>J</i>	0.62	0.31	0.29	0.63	0.59

components. This would increase the redundancy accounting capability of the PLS-PM estimates of the endogenous LVs and also match the formative scheme of the models underlying the data, as in the examples we have discussed in this paper.

Finally, we have performed a LISREL model analysis on each simulation case. Referring to Table 17, it can be seen that the goodness of fit is high in every situation, however the correlations between the two maximally different solutions of the same identified models are in many cases far from 1.

## 9. Conclusions and further research

The paper proposes a new approach for the estimation of a structural equation model with a formative-reflective scheme. The proposal is entitled RA-PM (redundancy analysis approach to path modeling) and consists of an alternative approach to the PLS-PM approach introduced by Lohmöller (1989).

The new RA-PM algorithm provides unique solutions for both the LVs and the parameters by preserving the causal structure of the model and utilizing all the relevant information in the formative MVs. The solutions are obtained by means of an iterative procedure which optimizes the redundancy criterion of Stewart and Love (1968).

Future developments will address with the extension of the methodology proposed to the case of categorical MVs based on the non-symmetric correspondence analysis quantification. In fact, it has been demonstrated that a relationship exists between redundancy analysis and non-symmetric correspondence analysis (Lauro and D'Ambra, 1984). As a consequence, it is possible to consider in the model both simple and interaction effects among categorical variables. Moreover, given that redundancy analysis represents a special case of PLS-PM with two sets of variables in the formative-reflective scheme (Tenenhaus et al., 2005), the iterative steps of PLS regression can be successfully introduced in the new proposed algorithm in order to overcome the problems related to missing data, multicollinearity and flat tables.

## Acknowledgements

The authors would like to express their thanks to the referees for their valuable remarks and comments.

## References

- Banca d'Italia, 2002. I bilanci delle famiglie italiane nell'anno 2000. Supplemento al Bollettino Statistico, Banca d'Italia, Roma.
- D'Ambra, L., Lauro, C.N., 1982. A principal component analysis with respect to a reference subspace. *Statist. Appl.* 15, 51–67.
- Guttman, L., 1955. The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British J. Math. Statist. Psych.* 8 (II), 65–81.
- Jöreskog, K.G., 1973. A general method for estimating a linear structural equation system. In: Goldberger, A.S., Duncan, O.D. (Eds.), *Structural Equation Models in the Social Sciences*. Seminar Press, New York, pp. 85–112.
- Jöreskog, K.G., 1981. Analysis of covariance structures. *Scand. J. Statist.* 8, 65–92.
- Jöreskog, K.G., 2000. Latent Variable Scores and Their Uses. (<http://www.ssicentral.com/lisrel/techdocs/lvscores.pdf>).

- Lauro, C.N., D'Ambra, L., 1984. Non-symmetrical correspondence analysis. In: Diday, E. et al. (Eds.), *Data Analysis and Informatics III*. North-Holland, Amsterdam, pp. 433–446.
- Lohmöller, J.B., 1989. *Latent Variable Path Modeling with Partial Least Squares*. Physica-Verlag, Heidelberg.
- Saris, W.E., Stronkhorst, L.H., 1984. *Causal Modeling in Non Experimental Research*. Sociometric Research Foundation, Amsterdam.
- Schönemann, P.H., 1971. The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika* 35, 21–30.
- Stewart, D.K., Love, W.A., 1968. A general canonical correlation index. *Psych. Bulletin* 70, 160–163.
- Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. *Comput. Statist. Data Anal.* 48, 159–205.
- Van den Wollenberg, A.L., 1977. Redundancy Analysis: An Alternative for Canonical Correlation Analysis. *Psychometrika* 42, 207–219.
- Vittadini, G., 1989. Indeterminacy problem in the LISREL model. *Multivariate Behavioral Res.* 24 (4), 397–414.
- Vittadini, G., Lovaglio, P.G., 2004. The estimate of human capital from two sets of observed indicators: formative and reflective. In: *Proceedings of the ASA Joint Statistical Meeting, Section on Business and Economic Statistics*. CD-ROM, Toronto, pp. 1362–1369.
- Wold, H., 1982. Soft modeling: the basic design and some extensions. In: Jöreskog, K.G., Wold, H. (Eds.), *Systems Under Indirect Observation: Causality, Structure, Prediction*. North-Holland, Amsterdam, pp. 1–54.