Structural Equation Modeling by Extended Redundancy Analysis[1]

Heungsun Hwang and Yoshio Takane[2]
Department of Psychology
McGill University
1205 Dr. Penfield Avenue, Montreal, Quebec, Canada H3A 1B1

Summary: A new approach to structural equation modeling, so-called extended redundancy analysis (ERA), is proposed. In ERA, latent variables are obtained as linear combinations of observed variables, and model parameters are estimated by minimizing a single least squares criterion. As such, it can avoid limitations of covariance structure analysis (e.g., stringent distributional assumptions, improper solutions, and factor score indeterminacy) in addition to those of partial least squares (e.g., the lack of a global optimization). Moreover, data transformation is readily incorporated in the method for analysis of categorical variables. An example is given for illustration.

1. Introduction

Two different approaches have been proposed for structural equation modeling (Anderson & Gerbing, 1988; Fornell & Bookstein, 1982). One analyzes covariance matrices as exemplified by covariance structure analysis (Jöreskog, 1970), while the other analyzes data matrices as exemplified by partial least squares (PLS, Wold, 1982). Typically covariance structure analysis estimates model parameters by the maximum likelihood method under the assumption of multivariate normality of variables. Yet, such a distributional assumption is often violated. A more serious problem is improper solutions (e.g., negative variance estimates), which occur with high frequency in practice. Also, factor scores or latent variable scores are indeterminate. An asymptotically distribution-free (ADF) estimator (Browne, 1984) can be used to fit non-normal data . The ADF estimation, however, is accurate only with very large samples and is still not free from improper solutions and factor score indeterminacy.

In PLS, on the other hand, latent variables are obtained as exact linear composites of observed variables and model parameters are estimated by the fixed-point algorithm (Wold, 1965). As such, PLS does not need any restrictive distributional assumptions. Moreover, PLS does not suffer from improper solutions and indeterminate factor scores. PLS, however, does not solve a global optimization problem for parameter estimation. The lack of a global optimization feature makes it difficult to evaluate an overall model fit. Also, it is not likely that the obtained PLS solutions are optimal in any well defined sense (Coolen & de Leeuw, 1987).

---

1. Published in S. Nishisato, Y. Baba, H. Bozdogan & K Kanefuji (Eds.), *Measurement and Multivariate Analysis* (pp. 115-124), Tokyo: Springer Verlag.

2. Both authors have contributed equally to the paper, and the authorship reflects the alphabetical order of the two authors.

In this paper, we propose a new method that avoids the major drawbacks of the conventional methods. It may be called extended redundancy analysis (ERA). In ERA, latent variables are estimated as linear combinations of observed variables, so that there are no improper solutions and non-unique factor scores. Also, it employs a global least squares (LS) criterion to estimate model parameters. Thus, it offers an overall model fit without recourse to the normality assumption.

2. Extended Redundancy Analysis

Let $\mathbf{Z}^{(1)}$ denote an n by p matrix consisting of observed endogenous variables. Let $\mathbf{Z}^{(2)}$ denote an n by q matrix consisting of observed exogenous variables. When an observed variable is exogenous as well as endogenous, it is included in both $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$. Assume that the columns of the matrices are mean centered and scaled to unit variance. Then, the model for extended redundancy analysis is given by

$$\mathbf{Z}^{(1)} = \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E} = \mathbf{F}\mathbf{A}' + \mathbf{E}, \qquad (2.1)$$

with

$$\text{rank}(\mathbf{W}\mathbf{A}') \leq \min(q,p),$$

where $\mathbf{W}$ is a matrix of weights, $\mathbf{A}'$ is a matrix of loadings, $\mathbf{E}$ is a matrix of residuals, and $\mathbf{F}$ (= $\mathbf{Z}^{(2)}\mathbf{W}$) is a matrix of component scores with identification restrictions $\text{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$. In (2.1), $\mathbf{W}$ and/or $\mathbf{A}'$ are structured according to the model to be fitted. Model (2.1) reduces to the redundancy analysis model (van den Wollenberg, 1977) when no variables are shared by both $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$, and no constraints other than $\text{rank}(\mathbf{W}\mathbf{A}')$ are imposed on $\mathbf{W}$ and $\mathbf{A}'$.

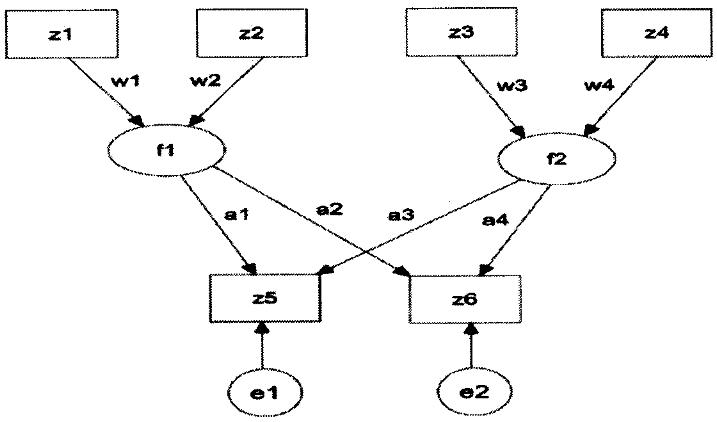

Figure 1. A two latent variable model among three sets of variables.

For simple illustration, suppose that there are three sets of variables, for example, $\mathbf{Z}_1 = [\mathbf{z}_1, \mathbf{z}_2]$, $\mathbf{Z}_2 = [\mathbf{z}_3, \mathbf{z}_4]$, and $\mathbf{Z}_3 = [\mathbf{z}_5, \mathbf{z}_6]$. Further suppose that there are relationships among the three sets of variables, as displayed in Figure 1. Figure 1 shows that two latent variables, one obtained from $\mathbf{Z}_1$ (i.e., $\mathbf{f}_1$), and the other from $\mathbf{Z}_2$ (i.e., $\mathbf{f}_2$), are combined to affect $\mathbf{Z}_3$.This may be expressed as

$$\mathbf{Z}_3 = [\mathbf{Z}_1 \,\vdots\, \mathbf{Z}_2]\begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ 0 & w_3 \\ 0 & w_4 \end{bmatrix}\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + \mathbf{E},$$

$$= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E} = \mathbf{F}\mathbf{A}' + \mathbf{E}, \tag{2.2}$$

where $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2]$, $\mathbf{W} = \begin{bmatrix} w_1 & w_2 & 0 & 0 \\ 0 & 0 & w_3 & w_4 \end{bmatrix}'$, $\mathbf{A}' = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$, and $\mathbf{F} = \mathbf{Z}^{(2)}\mathbf{W} = [\mathbf{f}_1 \,\vdots\, \mathbf{f}_2]$.

To estimate parameters, we seek to minimize the following LS criterion:

$$f = \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}') = \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{F}\mathbf{A}'), \tag{2.3}$$

with respect to $\mathbf{W}$ and $\mathbf{A}'$, subject to $\mathrm{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$, where $\mathrm{SS}(\mathbf{X}) = \mathrm{tr}(\mathbf{X}'\mathbf{X})$. An alternating least squares (ALS) algorithm is developed to minimize (2.3), which is a simple adaptation of Kiers and ten Berge (1989)'s algorithm.

To employ the ALS algorithm, we may rewrite (2.3) as

$$f = \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - \mathrm{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')) \tag{2.4a}$$

$$= \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - (\mathbf{A} \otimes \mathbf{Z}^{(2)})\mathrm{vec}(\mathbf{W})) \tag{2.4b}$$

$$= \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - (\mathbf{I} \otimes \mathbf{F})\mathrm{vec}(\mathbf{A}')) \tag{2.4c}$$

where $\mathrm{vec}(\mathbf{X})$ denotes a supervector formed by stacking all columns of $\mathbf{X}$, one below another, and $\otimes$ denotes a Kronecker product. The algorithm can then be made to repeat the following steps until convergence is reached.

(Step 1) Update $\mathbf{W}$ for fixed $\mathbf{A}'$ as follows: let $\mathbf{w}$ denote the vector formed by eliminating zero elements from $\mathrm{vec}(\mathbf{W})$ in (2.4b). Let $\mathbf{\Omega}$ denote the matrix formed by eliminating the columns of $\mathbf{A} \otimes \mathbf{Z}^{(2)}$ in (2.4b) corresponding to the zero elements in $\mathrm{vec}(\mathbf{W})$. Then, the LS estimate of $\mathbf{w}$ is obtained by

$$\tilde{\mathbf{w}} = (\mathbf{\Omega}'\mathbf{\Omega})^{-1}\mathbf{\Omega}'\mathrm{vec}(\mathbf{Z}^{(1)}). \tag{2.5}$$

The updated $\mathbf{W}$ is reconstructed from $\tilde{\mathbf{w}}$, and $\mathbf{F} = \mathbf{Z}^{(2)}\mathbf{W}$ is normalized so that $\mathrm{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$.

(**Step 2**) Update $\mathbf{A}'$ for fixed $\mathbf{W}$ as follows: let $\mathbf{a}$ denote the vector formed by eliminating zero elements from vec($\mathbf{A}'$) in (2.4c). Let $\boldsymbol{\Gamma}$ denote the matrix formed by eliminating the columns of $\mathbf{I} \otimes \mathbf{F}$ in (2.4c) corresponding to the zero elements in vec($\mathbf{A}'$). Then, the LS estimate of $\mathbf{a}$ is obtained by

$$\tilde{\mathbf{a}} = (\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\text{vec}(\mathbf{Z}^{(1)}). \tag{2.6}$$

The updated $\mathbf{A}'$ is recovered from $\tilde{\mathbf{a}}$.

In the method, the total fit of a hypothesized model to data is measured by the total variance of the observed endogenous variables explained by the exogenous variables. This is given by

$$\text{Fit} = 1 - \frac{\text{SS}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')}{\text{SS}(\mathbf{Z}^{(1)})}. \tag{2.7}$$

This fit index ranges from 0 to 1. The larger is the fit value, the more variance of the endogenous variables is explained by the exogenous variables. The standard errors of parameter estimates can be estimated by the bootstrap method (Efron, 1982). The bootstrapped standard errors can be used to assess the reliability of the parameter estimates. The critical ratios can be used to examine the significance of the parameter estimates (e.g., a parameter estimate having a critical ratio greater than two in absolute value is considered significant at a .05 significance level).

3. Analysis of categorical variables by data transformation

ERA can readily analyze categorical variables through a certain type of data transformation, often called optimal scaling (e.g., Young, 1981). In optimal scaling, the data are parametrized as $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, which are estimated, subject to constraints imposed by the measurement characteristics of $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$. We divide all parameters into two subsets: the model parameters and the data parameters. We then optimize a global fitting criterion by alternately updating one subset with the other fixed. Note that $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ may contain variables with different measurement characteristics. This means that a variable may not be directly comparable with other variables, so that each variable in $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ should be separately updated.

The ALS procedure with the data transformation feature proceeds as follows. Let $\mathbf{z}_i$ denote a variable in either $\mathbf{Z}^{(1)}$ or $\mathbf{Z}^{(2)}$, so that $i = 1, \cdots, p + q$. Let $\mathbf{s}_i$ denote a variable in either $\mathbf{S}^{(1)}$ or $\mathbf{S}^{(2)}$. Then, we seek to minimize

$$f = \text{SS}(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\mathbf{W}\mathbf{A}') = \text{SS}(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\mathbf{B}), \tag{3.1}$$

with the conditions that $\text{diag}(\mathbf{W}'\mathbf{S}^{(2)\prime}\mathbf{S}^{(2)}\mathbf{W}) = \mathbf{I}$, $\mathbf{s}_i'\mathbf{s}_i = 1$, and $\mathbf{s}_i = \xi(\mathbf{z}_i)$, where $\mathbf{B} = \mathbf{W}\mathbf{A}'$, and ξ refers to a transformation of the observations in $\mathbf{z}_i$, which is a function of their measurement characteristics. To minimize (3.1), two main phases are alternated. One phase is the model estimation phase, in which the model parameters are estimated. The other is the data transformation phase that estimates the data parameters. The model estimation phase is analogous to the estimation procedure in Section 2. We thus focus on the data transformation phase here. The data transformation phase mainly consists of two steps. In the first step, the model prediction of $\mathbf{s}_i$ is obtained in such a way that it minimizes (3.1). In the next step, $\mathbf{s}_i$ is transformed in such a way that it maximizes the relationship between $\mathbf{s}_i$ and the model prediction under certain measurement restrictions.

The first step of the data transformation phase is given as follows. Let $\mathbf{s}_g^{(1)}$ and $\mathbf{s}_h^{(2)}$ denote the g-th and h-th variables in $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, respectively ($g = 1, \cdots, p$; $h = 1, \cdots, q$). Let $\tilde{\mathbf{s}}_i$ denote the model prediction of $\mathbf{s}_i$. Then (3.1) may be rewritten as

$$f = \sum_{i=1}^{p+q} \mathrm{SS}(\mathbf{s}_i \boldsymbol{\eta}' - (\boldsymbol{\Delta} - \boldsymbol{\Psi})). \tag{3.2}$$

In (3.2), $\boldsymbol{\eta}'$, $\boldsymbol{\Delta}$, and $\boldsymbol{\Psi}$ are defined as follows: Suppose that if $\mathbf{s}_i$ is shared by $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, it is placed in the g-th column and the h-th column of $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, respectively. Then, when the model predictions of the variables in $\mathbf{S}^{(1)}$ are updated,

$$\boldsymbol{\Delta} = \begin{cases} \mathbf{S}_{(h)}^{(2)}\mathbf{B}_{(h)} & \text{if } \mathbf{s}_i \text{ is shared} \\ \mathbf{S}^{(2)}\mathbf{B} & \text{otherwise} \end{cases}, \quad \boldsymbol{\Psi} = \mathbf{S}_{(g)}^{(1)}, \text{ and } \quad \boldsymbol{\eta}' = \begin{cases} \mathbf{e}_g' - \mathbf{b}_h' & \text{if } \mathbf{s}_i \text{ is shared} \\ \mathbf{e}_g' & \text{otherwise} \end{cases}.$$

When the model predictions of non-common variables in $\mathbf{S}^{(2)}$ are updated,

$$\boldsymbol{\Delta} = \mathbf{S}_{(h)}^{(2)}\mathbf{B}_{(h)}, \quad \boldsymbol{\Psi} = \mathbf{S}^{(1)}, \quad \text{and } \boldsymbol{\eta}' = \mathbf{b}_h'.$$

In the above, matrix $\mathbf{S}_{(h)}^{(2)}\mathbf{B}_{(h)}$ is a product of $\mathbf{S}^{(2)}$ whose h-th column is the n-component vector of zeros and $\mathbf{B}$ whose h-th row is the p-component vector of zeros. Matrix $\mathbf{S}_{(g)}^{(1)}$ equals to $\mathbf{S}^{(2)}$ whose g-th column is an n-component vector of zeros. $\mathbf{e}_g'$ denotes a p-component row vector whose elements are all zeros except the g-th element being unity. Vector $\mathbf{b}_h'$ corresponds with the h-th row of $\mathbf{B}$. Then, $\tilde{\mathbf{s}}_i$ is obtained by

$$\tilde{\mathbf{s}}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}(\boldsymbol{\eta}'\boldsymbol{\eta})^{-1}, \tag{3.3}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Delta} - \boldsymbol{\Psi}$.

In the next step, $\mathbf{s}_i$ is transformed in such a way that it is close to $\tilde{\mathbf{s}}_i$ as much as possible under the appropriate measurement restrictions. In many cases, $\mathbf{s}_i$ is updated by minimizing a LS fitting criterion (e.g., the (normalized) residuals between $\mathbf{s}_i$ and $\tilde{\mathbf{s}}_i$). This comes down to regressing $\tilde{\mathbf{s}}_i$ onto the space of $\mathbf{z}_i$, which represents the measurement restrictions. The LS estimate of $\mathbf{s}_i$ can be generally expressed as follows

$$\mathbf{s}_i = \boldsymbol{\Upsilon}_i(\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i)^{-1}\boldsymbol{\Upsilon}_i'\tilde{\mathbf{s}}_i. \tag{3.4}$$

In (3.4), $\boldsymbol{\Upsilon}_i$ is determined by the measurement restrictions imposed on the transformation. For example, for nominal variables, $\boldsymbol{\Upsilon}_i$ is an indicator matrix, whose element stands for category membership, and is known in advance. For ordinal variables, Υ_i indicates which categories must be blocked to satisfy the ordinal restriction, and is iteratively constructed by Kruskal's (1964) least squares monotonic transformation algorithm. The updated $\mathbf{s}_i$ is then normalized to satisfy $\mathbf{s}_i'\mathbf{s}_i = 1$. In (3.2), we see that updating a variable is dependent on other variables. To ensure convergence, we must immediately replace the previously estimated variable by the newly estimated and normalized variable. Moreover, when $\mathbf{s}_i$ is included in both $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, the rescaled and normalized $\mathbf{s}_i$ should be substituted for the corresponding columns in both $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$.

4. Example

The present example is part of the so-called basic health indicator data collected by the World Health Organization. They are available through the internet (http://www.who.int). It consists of 6 variables measured in different countries: (1) infant mortality rate (IMR), defined as the number of deaths per 1000 live births between birth and exact age one year in 1998. (2) maternal mortality ratio (MMR), defined as the number of maternal deaths per 100000 live births in 1990, (3) real gross domestic product (GDP) per capita adjusted for purchasing power parity is expressed in 1985 US dollars, (4) the average number of years of education given for females aged 25 years and above (FEUD) (5) the percentage of children immunized against measles in 1997 (Measles), and (6) total health expenditures as a percentage of GDP in 1995 (Healthexp). The sample size is 51, which amounts to the number of countries for which the data are available.

Two latent variables were assumed for the last four observed variables. One latent variable called 'social and economic (SE) factor' was defined as a linear combination of GDP and FEUD, and the other called 'health services (HS) factor' as that of Measles and Healthexp. The two latent variables were in turn deemed to influence two observed endogenous variables, IMR and MMR. For this model, $\mathbf{W}$ and $\mathbf{A}'$ were identical to those in (2.2). By using ERA, the model was fitted to the data. Results are provided in Figure 2.

The goodness of fit of the model was .65, indicating that about 65% of the variance of the endogenous variables were accounted for by the two latent variable model. The fit turned out to be significant in terms of its critical ratio obtained from the bootstrap method with 100 bootstrap samples, indicating that the fitted model was significantly different from the model which assumed $\mathbf{B} = \mathbf{0}$. The squared multiple correlations of IMR and MMR were .73 and .57, respectively. They also turned out significant according to their bootstrapped critical ratios. In Figure 2, boldfaced parameter estimates indicate that they turned out to be significant in terms of their critical ratios. The component weights associated with SE were all significant and negative. This indicates that SE was characterized as social and economic underdevelopment. Similarly, the component weights of Mealses and Healthexp were negative, indicating that HS was likely to represent a low level of health services.

However, only one variable, Measles, was significantly associated with HS. Both latent variables were found to have a significant and positive effect on IMR and MMR. This indicates that social and economic underdevelopment and the low level of health services are likely to increase infant mortality rate and maternal mortality ratio. The correlation between the two latent variables was .47.

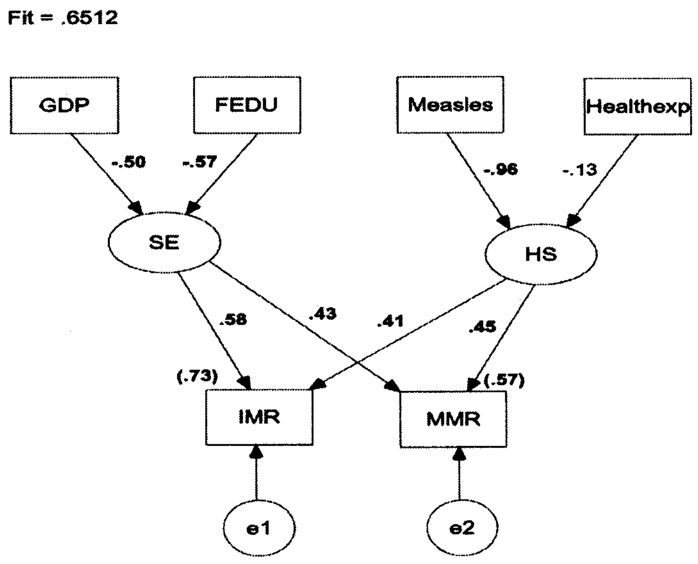

Figure 2. Results of fitting the two latent variable model for the WHO data.

To exemplify data transformations, two observed endogenous variables, that is, IMR and MMR, were monotonically transformed. Kruskal's (1964) primary LS monotonic transformation was applied to them. This indicated that observation categories were order-preserved but tied observations might become untied. The LS monotonic transformations of the variables are shown in Figures 3 and 4. The left-hand and right-hand figures represent the LS monotonic transformations of IMR and MMR, respectively. In both figures, the original observations (horizontal) are plotted against the transformed scores (vertical). We find that the monotonic transformations are quite steep although they contain some ties. Due to the transformation, the fit of the model was dramatically improved (.96), while providing similar interpretations of parameter estimates as those obtained when the variables were treated as numerical.

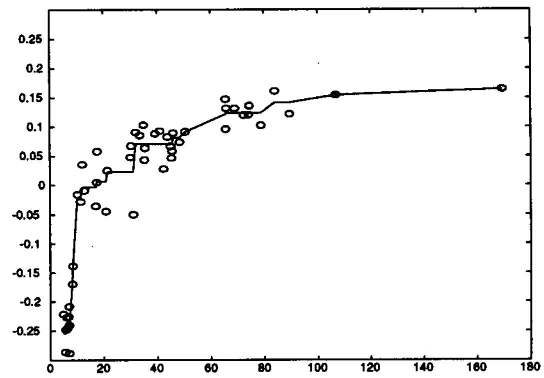

Figure 3. The LS monotonic transformation of IMR

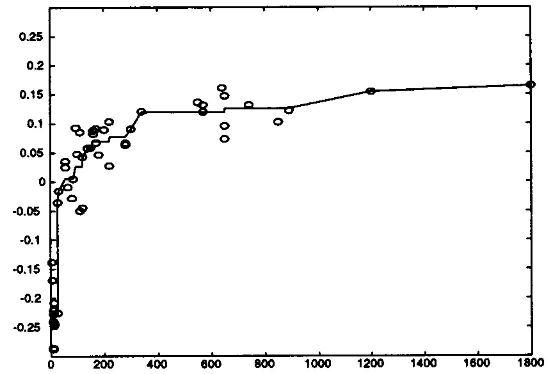

Figure 4. The LS monotonic transformation of MMR.

5. Concluding Remarks

A few attempts have been made to extend redundancy analysis to three sets of variables (e.g., Takane, Kiers, & de Leeuw, 1995; Velu, 1991). Yet, they are limited to model and fit a particular type of relationship among three sets of variables. Our method, on the other hand, is quite comprehensive in extending redundancy analysis, and it enables us to specify and fit various structural equation models. Although they are not presented here to conserve space, the basic ERA model can be readily extended to fit more complex relationships among variables, including direct effects of observed variables and higher-order latent variables. Moreover, it is able to perform multi-sample comparisons (Takane & Hwang, 2000).

The data transformation may be considered as one of the principal assets of our method. This makes the data more in line with the model, and goodness of fit may be improved. This also allows us to examine relationships among various types of data measured at different levels. This kind of data transformation is feasible because our method directly analyzes the data matrices rather than the covariance or correlation matrix. However, in PLS, which also analyzes the data matrices, this particular way of data transformation is not feasible since it requires a well-defined global criterion that is consistently optimized by updating the transformed variables.

A number of relevant topics may be considered to further enhance the capability of the method (Hwang & Takane, 2000). For instance, robust estimation may be in order since the proposed method may not be robust against outliers as far as it is based on solving a simple (unweighted) least squares criterion. Missing observations can raise a serious problem, which frequently appear in large data sets. The assumption of normality is not essential for the method due to the least squares fitting. If it is assumed, nonetheless, we can extend the current estimation method in such a way that it provides efficient estimators and allows to perform statistical significance tests without recourse to resampling methods. Future studies are needed to deal with such topics in the proposed method.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, **103**, 411-423.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Coolen, H., & de Leeuw, J. (1987). Least squares path analysis with optimal scaling, Paper presented at the fifth international symposium of data analysis and informatics. Versailles, France.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: SIAM.

Hwang, H., & Takane, Y. (2000). Structural equation modeling by extended redundancy analysis: Extended features. Manuscript submitted for publication.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, **57**, 409-426.

Kiers, H. A. L., & ten Berge, J. M. F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in two or more populations. *Psychometrika*, **54**, 467-473.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.

Takane, Y., & Hwang, H. (2000). Structural equation modeling by extended redundancy analysis: Basic features. Manuscript submitted for publication.

Takane, Y., Kiers, H., & de Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika*, **60**, 259-280.

van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical analysis. *Psychometrika*, **42**, 207-219.

Velu, R. P. (1991). Reduced rank models with two sets of regressors. *Applied Statistics*, **40**, 159-170.

Wold, H. (1965). A fixed-point theorem with econometric background, I-II. *Arkiv for Matematik*, **6**, 209-240.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog and H. Wold (Eds.), *Systems Under Indirect Observations: Causality, Structure, Prediction, Vol.* 2 (pp. 1-54). Amsterdam: North-Holland.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, **46**, 357-388.