INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Structural equation modeling by extended redundancy analysis

Heungsun Hwang
Department of Psychology
McGill University
Montreal

November 2000

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Table of Contents

Abstract

A new approach to structural equation modeling based on so-called extended redundancy analysis (ERA) is proposed. In ERA, latent variables are obtained as exact linear combinations of observed variables, and model parameters are estimated by consistently minimizing a single criterion. As a result, the method 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 procedure). The method is simple yet versatile enough to fit more complex models; e.g., those with higher-order latent variables and direct effects of observed variables. It can also fit a model to more than one sample simultaneously. Other relevant topics are also discussed, including data transformations, missing data, metric matrices, robust estimation, and efficient estimation. Examples are given to illustrate the proposed method.

Résumé

Nous proposons une nouvelle approche de modélisation des équations structurelles basée sur une analyse des redondances étendue. Dans ce type d'analyse, les variables latentes sont obtenues par combinaisons linéaires exactes de variables observées et les paramètres du modèle sont estimés en minimisant un critère unique. Cette méthode permet d'éviter les limites asssociées à l'analyse de structure de covariance - tels que les postulats relatifs à la distribution des variables, les solutions impropres et l'indétermination de scores factoriels - de même que celles liées à la méthode des moindres carrés partiels - comme l'absence de procédure d'optimisation globale. La méthode est simple tout en étant assez flexible pour ajuster des modèles aussi complexes que ceux comprenant des variables latentes d'ordre supérieur et des effets directs de variables observées. Elle permet également l'ajustement d'un modèle, de facon simultanée, à des données issues d'échantillons différents. Nous aborderons également d'autres sujets pertinents, tels que la transformation des données, le traitement de données manquantes, la construction de matrices métriques, la robustesse et l'efficacité de l'estimation. Des exemples illustrent l'application de la méthode.

Acknowledgments

I wish to express my gratitude to many people who contributed to the completion of this dissertation. Foremost, I would like to thank my supervisor, Yoshio Takane, for his generous support and guidance during the years I worked at McGill Psychology. He gave me the opportunity to be his very first student in a twenty-year academic career, and led me to the area of quantitative psychology. He was always there to answer my questions big or small. He is my most admired mentor. Yoshio taught me what a scholar is as well as quantitative psychology itself.

I am indebted to Jim Ramsay for his help in smooth monotonic data transformations, which are one of the key features in my dissertation. Moreover, he and his son, Tim, inspired me to realize that philosophy is closely connected with statistics. I would also like to thank David Zuroff for his helpful comments on this research. Special thanks to Nicole Malfait for the French translation of the abstract.

Finally, I am grateful to my family in Montreal and Seoul for their support and patience. Especially, I would like to thank my wife, Seungmi. This endeavour would not have been possible without her unwavering support and faith.

Statement of Original Contributions

The present research develops a new statistical method for analysis of structural equation models. The method overcomes major limitations of conventional methods (e.g., the occurrence of improper solutions in covariance structure analysis, and the lack of a global optimization criterion in partial least squares). The method allows to fit diverse complex relationships among variables, including direct effects of observed variables, higher-order latent variables, and multi-sample comparisons. In addition, it can handle data transformations, missing data, robust estimation, and efficient estimation in a simple way.

Chapter 1

Introduction

Structural equation models are used to specify and test hypothesized relationships among observed variables and unobserved variables (called latent variables) in multivariate data. Structural equation models allow us to specify simultaneous equations among multiple sets of independent variables and dependent variables. They enable us to construct more general models than traditional factor or principal component analytic procedures. They also allow for statistical tests to assess and modify theoretical models, providing insight for further model generation (Jöreskog, 1993). That is, structural equation models involve generalizations and extensions of traditional multivariate analysis techniques.

Two different approaches have been proposed for structural equation models (Anderson & Gerbing, 1988; Fornell & Bookstein, 1982; Jöreskog & Wold, 1982): Covariance structure analysis and partial least squares. The former analyzes covariance matrices derived from data matrices, whereas the latter directly analyzes data matrices. In covariance structure analysis (Bock & Bargmann, 1966; Jöreskog, 1970, 1973, 1977), the structure of the population covariance matrix is modeled as a function of parameters of the specified structural equation model. The modeled population covariance matrix is often called the implied population covariance matrix (Bollen,

1989). If the model is correct and the population covariance matrix is known, we can estimate the parameters by minimizing the difference between the population covariance matrix and the implied population covariance matrix. In practice, however, the population covariance matrix is unknown. Then its consistent estimators, the sample covariance matrix, has to be used to estimate the parameters. Under the assumption of multinormality, Jöreskog (1970, 1973) developed a maximum likelihood (ML) estimation method for fitting the covariance structure analysis model. The ML estimation method is by far the most widely used method (Bollen, 1989), although the generalized least squares (GLS) and unweighted least squares (ULS) estimation methods (Browne, 1982, 1984) are also employed. Much of the theoretical background for covariance structure analysis can be found in Bollen (1989), Hayduk (1987), Hoyle (1995), Kline (1998), Loehlin (1998), and so on.

Partial least squares (PLS), on the other hand, introduced by Herman Wold (1966, 1973, 1975, 1982), estimates latent variables as exact linear combinations of observed variables (this is also called the 'weighted relations'). In PLS, model parameters are estimated by a fixed point (FP) algorithm (Lyttkens, 1968, 1973; Wold, 1965, 1981). In the FP algorithm used in PLS, a set of model parameters are divided into subsets, and each subset is 'partially' estimated by ordinary multiple regression analysis with other subsets fixed. This partial estimation is cycled through repeatedly until convergence is reached. For more recent information on PLS, refer to Falk and Miller (1992), Fornell and Cha (1994), Lohmöller (1989), and so on.

Covariance structure analysis and PLS deal with structural equation models from

statistically different perspectives. Velicer and Jackson (1990) pointed out that the former entails factor analysis, whereas the latter involves component analysis (Meredith & Millsap, 1985; Schönemann & Steiger, 1976). That is, the latent variables in covariance structure analysis are equivalent to common factors while those in PLS are equivalent to components. Therefore, both techniques have their relative advantages and disadvantages. Covariance structure analysis seeks overall optimization in parameter estimates through a full information estimation technique (ML or GLS). As such, it provides efficient and consistent parameter estimates (Jöreskog & Wold, 1982). It also offers a variety of parametric tests for an overall model fit. Nonetheless, the distributional assumptions required for covariance structure analysis (e.g., multinormality of the observed variables) are often violated (Micceri, 1989). Relatively large sample size, say more than 100 (Boomsma, 1985), is also recommended to validate its use. The problem of improper solutions (e.g., factor correlation estimates greater than ± 1, negative variance estimates, etc.) is even more serious (Fornell & Bookstein, 1982). The improper solutions make it difficult to interpret the obtained results, but unfortunately, they occur with high frequency. In addition, factor scores or latent variable scores are indeterminate, which indicates that we can calculate different factor scores to fit the model equally well (e.g., McDonald & Mulaik, 1979; Steiger, 1979).

The normality assumption can be dropped by utilizing asymptotically distribution-free (ADF) estimators (e.g., Browne, 1982, 1984; Meijer, 1998). The ADF estimation, however, is more computationally expensive than the maximum likelihood

estimation, and is accurate only with very large samples (e.g., Muthén & Kaplan, 1992). The factor score indeterminacy problem seems to have minor practical consequences in the light of high degrees of similarity between factor and component solutions (Velicer & Jackson, 1990). On the other hand, it seems that there exists no obvious remedy against improper solutions in covariance structure analysis. Although it has been suggested that an improper solution could be considered as a diagnostic for model mispecification (e.g., Jöreskog & Sörbom, 1989, p. 239), it may occur even when the correct model is specified (Kiers, Takane, & ten Berge, 1996).

PLS does not need any stringent distributional assumptions because its model parameters are estimated on the basis of partial OLS. Moreover, PLS does not suffer from improper solutions and indeterminate factor scores, since the latent variables are given by linear combinations of the observed variables. PLS, however, does not solve a global optimization problem for parameter estimation (Jöreskog & Wold, 1982; Fornell & Bookstein, 1982). This indicates that there exists no criterion consistently minimized or maximized to determine estimates of model parameters. The lack of a global optimization criterion makes it difficult to evaluate the PLS procedures (McDonald, 1996). More seriously, PLS has no mechanism to evaluate the overall fit of the model. It is not likely that the obtained PLS solutions are optimal in an overall fit (Coolen & de Leeuw, 1987).

Despite a number of benefits of both techniques for fitting structural equation models, in covariance structure analysis the occurrence of improper solutions is most likely to interfere with meaningful analysis, whereas in PLS the lack of a global

optimization criterion seems to make its use limited.

In this dissertation, I propose a new method that avoids the major drawbacks of the conventional methods, that is, improper solutions in covariance structure analysis and the lack of an overall optimization criterion in PLS. The proposed method will be called extended redundancy analysis. In short, extended redundancy analysis may be described as a kind of structured component analysis. In this method, latent variables are estimated as linear combinations of observed variables to avoid the problem of improper solutions as in PLS. Furthermore, it provides a global fitting criterion, which is consistently minimized to estimate parameters, to overcome the limitation of PLS.

The remaining chapters of this dissertation are organized as follow. In Chapter 2, the proposed method is studied in detail. First of all, redundancy analysis is briefly discussed, which is the prototype of the proposed method. The basic model for extended redundancy analysis and its parameter estimation are then presented. How to incorporate additional constraints into model parameters is also discussed. Examples are given to illustrate the feasibility of the proposed method. In Chapter 3, some possible extensions of the basic model are discussed. They include higher-order latent variables, direct effects of observed variables, and multi-sample comparisons. Each extension is empirically illustrated. Chapter 4 deals with other relevant topics to the proposed method, such as data transformations, missing data, metric matrices, and robust estimation. Examples are also presented for illustration. The proposed method can be extended, so that it can provide efficient estimators when the normality assumption is satisfied. The final chapter briefly summarizes the previous chapters and

discusses further prospects of the proposed method.

Chapter 2

Extended Redundancy Analysis

Extended redundancy analysis (ERA), as the name suggests, is an extension of redundancy analysis (van den Wollenberg, 1977). We therefore start this chapter with a brief description of redundancy analysis. We then discuss the extension (ERA) in detail, followed by parameter estimation, additional constraints on parameters, and illustrative examples.

2. 1. Redundancy Analysis

Redundancy analysis (RA) is a useful technique for finding a directional relationship between two sets of multivariate data, mediated by latent variables (Lambert, Wildt, & Durand, 1988). Technically redundancy analysis amounts to extracting a series of linear combinations or components from one set of observed variables in such a way that they are mutually orthogonal and successively explain the maximum variance of the other set of variables. Let $\mathbf{Z}_1$ denote an n by r matrix of dependent or endogenous variables. Let $\mathbf{Z}_2$ denote an n by t matrix of independent or exogenous variables. Then, the model for redundancy analysis may be written as

$$\begin{aligned} \mathbf{Z}_1 &= \mathbf{Z}_2\mathbf{W}\mathbf{A}' + \mathbf{E}, \\ &= \mathbf{F}\mathbf{A}' + \mathbf{E}, \end{aligned} \tag{2.1}$$

$$\text{rank}(\mathbf{WA}') = d \leq \kappa, \tag{2.2}$$

where $\mathbf{W}$ is a t by d matrix of component weights, $\mathbf{A}'$ is a d by r matrix of loadings, $\mathbf{E}$ is the residual matrix, $\mathbf{F}$ (= $\mathbf{Z}_2\mathbf{W}$) denotes the matrix of linear components, and $\kappa \leq \min(r,t)$. Due to the restriction on rank in (2.2), the above model is also called the reduced rank regression model (e.g., Anderson, 1951; Rao, 1964; Izenman, 1975; Davies & Tso, 1982; Reinsel & Velu, 1998).

To illustrate further, the redundancy analysis model may be expressed in terms of a path diagram, a pictorial representation of a system of equations, as given in Figure 2.1. In the figure, square boxes are used to indicate observed variables, circles are used to represent latent variables, and straight arrows are used to signify that the variable at the base of an arrow affects the variable at the head of the arrow. The path diagram indicates that the redundancy analysis model can be viewed as a simple type of structural equation model between two sets of variables, in which latent variables are specified as linear combinations of $\mathbf{Z}_2$ (i.e., $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_d] = \mathbf{Z}_2\mathbf{W}$), which influence $\mathbf{Z}_1$. This simple structural equation model is often called the multiple indicator/multiple causes (MIMIC) model (e.g., Jöreskog & Goldberger, 1975; Fornell, Barclay, & Rhee, 1988).

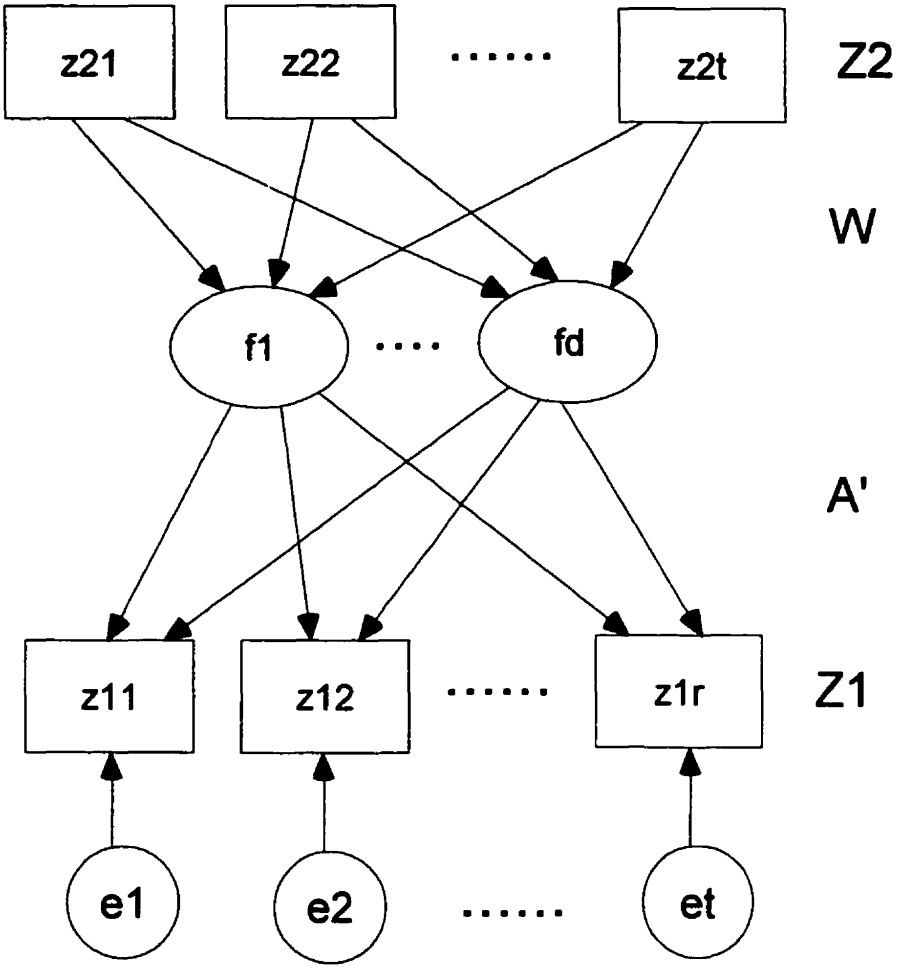

Figure 2.1. A path diagram for the redundancy analysis model.

To estimate parameters in redundancy analysis, we seek to minimize the following least squares (LS) criterion

$$f = \mathrm{SS}(\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{W}\mathbf{A}'), \tag{2.3}$$

where $\mathrm{SS}(\mathbf{X}) = \mathrm{trace}(\mathbf{X}'\mathbf{X})$, with respect to $\mathbf{W}$ and $\mathbf{A}'$. Here it is assumed that $\mathbf{W}'\mathbf{Z}_2'\mathbf{Z}_2\mathbf{W} = \mathbf{I}$ for identification. Minimizing (2.3) computationally comes down to calculating the generalized singular value decomposition (GSVD) of $(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2'\mathbf{Z}_1$ with metric matrices $\mathbf{Z}_2'\mathbf{Z}_2$ and $\mathbf{I}$ (e.g., Takane & Shibayama, 1991). (For the computation of GSVD, refer to Greenacre, 1984, Appendix A.) This indicates that ordinary redundancy analysis between two data sets has an analytic solution.

2.2. The Extended Redundancy Analysis Model

Given that the redundancy analysis model is a kind of structural equation model between two sets of variables, we may specify and fit a more variety of structural equation models by extending redundancy analysis to more than two sets of variables. For simple illustration, we 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 $\mathbf{Z}_3 = [\mathbf{z}_5, \mathbf{z}_6]$. We further suppose that there are relationships among the three sets of variables, as displayed in Figure 2.2.

Figure 2.2 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 relationship may be expressed as

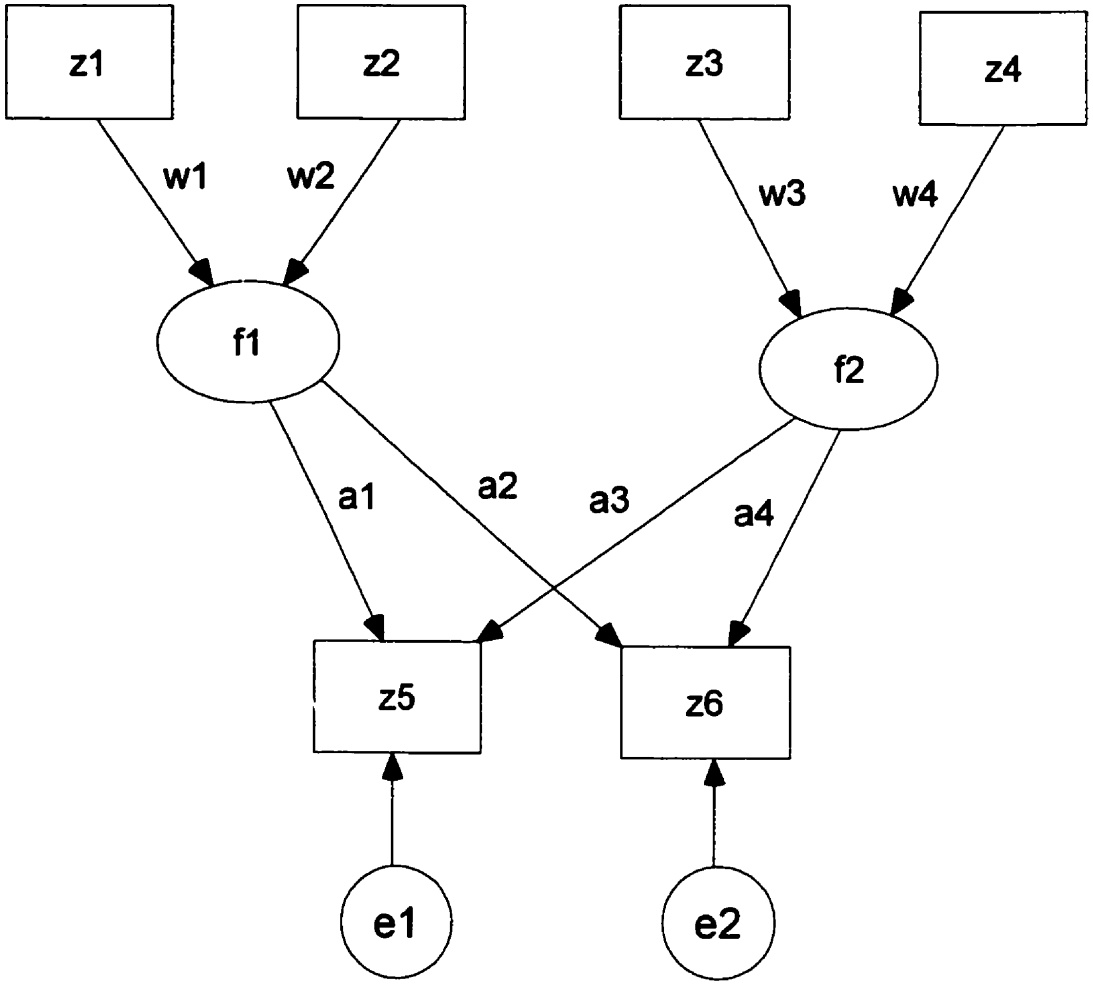

Figure 2.2. A path diagram for a model among three data sets (example 1).

$$\mathbf{Z}_3 = [\mathbf{Z}_1 \vdots \mathbf{Z}_2]\begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ 0 & w_3 \\ 0 & w_4 \end{bmatrix}\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + \mathbf{E}$$
$$= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E}$$
$$= \mathbf{F}\mathbf{A}' + \mathbf{E}, \qquad (2.4)$$

where $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2]$, $\mathbf{W} = \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ 0 & w_3 \\ 0 & w_4 \end{bmatrix}$, $\mathbf{A}' = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$, and $\mathbf{F} =$ $\mathbf{Z}^{(2)}\mathbf{W} = [\mathbf{f}_1 \vdots \mathbf{f}_2]$. Model (2.4) is essentially the same as (2.1). The difference is in that in (2.4) some elements of $\mathbf{W}$ are fixed as zeros, and two subvectors in $\mathbf{F}$, $\mathbf{f}_1$ and $\mathbf{f}_2$, are separately constrained to be of unit length.

We may consider another relationship among the three sets of variables, as presented in Figure 2.3. In Figure 2.3, we see that $\mathbf{f}_1$ has an effect on $\mathbf{Z}_2$ in addition to the effect on $\mathbf{Z}_3$. This relationship can be expressed as

$$[\mathbf{Z}_2 \vdots \mathbf{Z}_3] = [\mathbf{Z}_1 \vdots \mathbf{Z}_2]\begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ 0 & w_3 \\ 0 & w_4 \end{bmatrix}\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & a_5 & a_6 \end{bmatrix} + \mathbf{E}$$
$$= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E}$$
$$= \mathbf{F}\mathbf{A}' + \mathbf{E}, \qquad (2.5)$$

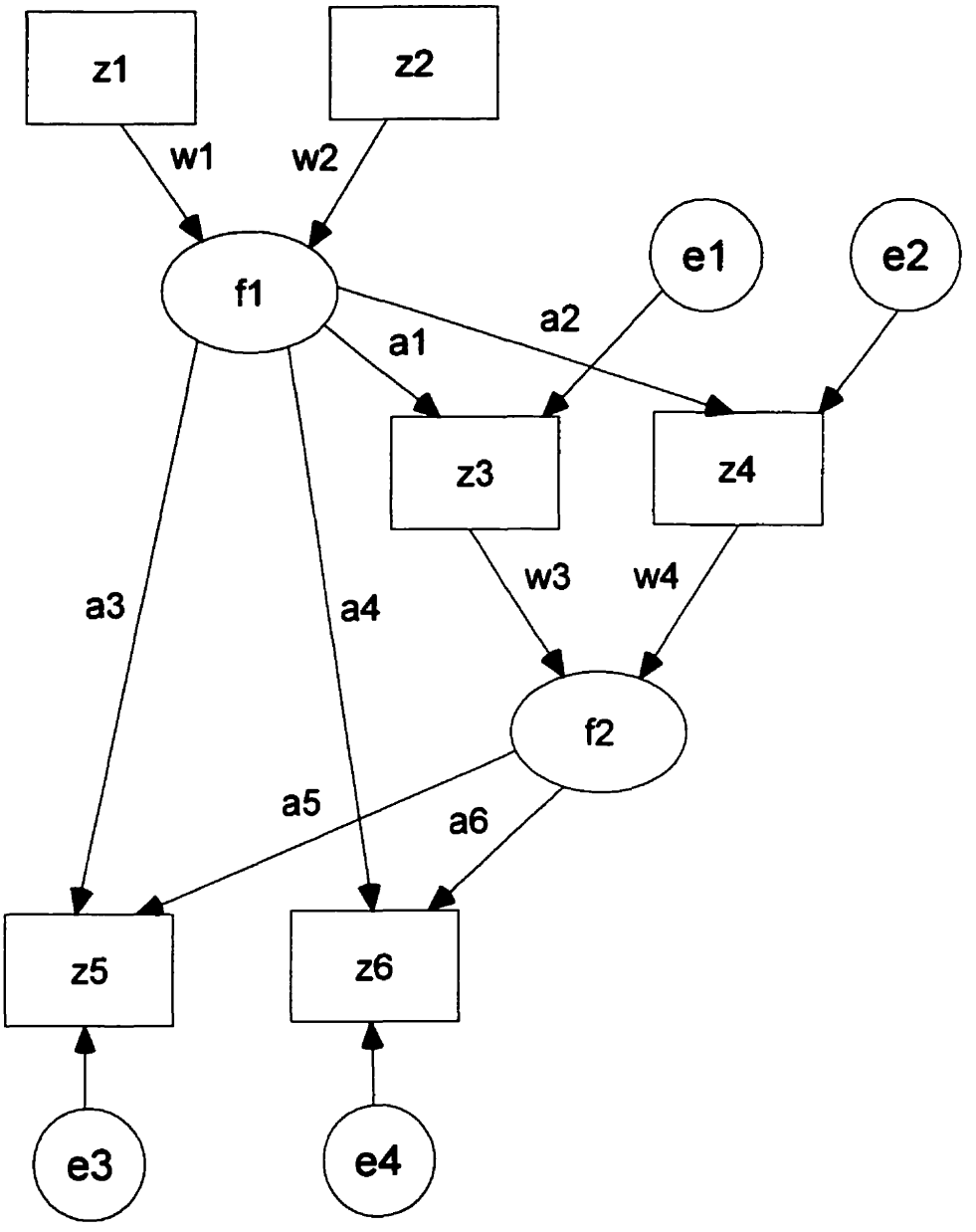

Figure 2.3. A path diagram for a model among three data sets (example 2).

where $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4]$, $\mathbf{W} = \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ 0 & w_3 \\ 0 & w_4 \end{bmatrix}$, and $\mathbf{A}' = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & a_5 & a_6 \end{bmatrix}$.

Again (2.5) is essentially of the same form as (2.1). However, both $\mathbf{W}$ and $\mathbf{A}'$ contain some fixed zero elements.

It may be seen that from the above examples the model for extended redundancy analysis (ERA) can generally be stated as follows: 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 can be generally expressed as

$$\begin{aligned} \mathbf{Z}^{(1)} &= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E} \\ &= \mathbf{F}\mathbf{A}' + \mathbf{E}, \end{aligned} \tag{2.6}$$

$$\text{rank}(\mathbf{W}\mathbf{A}') = d \leq v, \tag{2.7}$$

where $\mathbf{W}$ denotes a q by d matrix of component weights, $\mathbf{A}'$ denotes a d by p matrix of component loadings, $\mathbf{E}$ denotes an n by p matrix of residuals, $\mathbf{F}$ (= $\mathbf{Z}^{(2)}\mathbf{W}$) denotes an n by p matrix of component or latent variable scores, and $v \leq \min(q,p)$. For identification, $\mathbf{F}$ is restricted to be $\text{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$. As shown in the examples, $\mathbf{W}$ and/or $\mathbf{A}'$ in (2.6) are structured according to the model to be fitted.

Model (2.6) reduces to the redundancy analysis model when no variables are

shared by both $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$, and no constraints other than rank($\mathbf{WA}'$) are imposed on $\mathbf{W}$ and $\mathbf{A}'$. In (2.4) and (2.5), only a single linear component is extracted from $\mathbf{Z}_1$ and $\mathbf{Z}_2$ each, that is, $\mathbf{f}_1$ and $\mathbf{f}_2$, respectively. If more than one component are obtained from each set, we simply replace the vectors of component weights, say $\mathbf{w}_1 = [w_1, w_2]'$ and $\mathbf{w}_2 = [w_3, w_4]'$, by matrices of weights, imposing mutual orthogonality of the components.

A few attempts have been made to extend redundancy analysis to three sets of variables (e.g., Velu, 1991; Reinsel & Velu, 1998). However, they are limited to model and fit a particular type of relationship among three sets of variables. Model (2.6), on the other hand, is quite comprehensive in extending redundancy analysis, and it enables us to specify and fit various structural equation models.

2. 3. Parameter Estimation

We estimate the unknown parameters in model (2.6), $\mathbf{W}$ and $\mathbf{A}'$, in such a way that the sum of squares of the residuals $\mathbf{E} = \mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{WA}'$ is as small as possible. This amounts to minimizing

$$\begin{aligned} f &= \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{WA}') \\ &= \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{FA}'), \end{aligned} \tag{2.8}$$

with respect to $\mathbf{W}$ and $\mathbf{A}'$, subject to $\mathrm{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$.

Criterion (2.8) is essentially of the same form as redundancy analysis. Unlike redundancy analysis, however, minimizing (2.8) does not reduce to GSVD due to the structure in $\mathbf{W}$ and $\mathbf{A}'$. Instead, (2.8) should be minimized by an iterative method. We

use an alternating least squares (ALS) algorithm to minimize (2.8). Our algorithm is a simple adaptation of the ALS algorithm developed by Kiers and ten Berge (1989), that was used for simultaneous components analysis for two or more populations (Millsap & Meredith, 1988). In the algorithm, parameter matrices $\mathbf{W}$ and $\mathbf{A}'$ are alternately updated until convergence is reached. Updates of one parameter matrix are obtained in such a way that they minimize (2.8), with the other set fixed. The ALS algorithm is monotonically convergent (e.g., de Leeuw, Young, & Takane, 1976). This means that the function value will never increase throughout the iterations. Kiers and ten Berge's algorithm reduces to the algorithm developed here when $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ consist of distinct variables in a single population.

To employ the ALS algorithm, we may rewrite (2.8) as

$$f = \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - \mathrm{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')), \tag{2.9}$$

where vec($\mathbf{X}$) denotes a supervector consisting of all columns of $\mathbf{X}$, one below another. In (2.9),

$$\mathrm{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}') = (\mathbf{A} \otimes \mathbf{Z}^{(2)})\mathrm{vec}(\mathbf{W}) \tag{2.10a}$$

$$= (\mathbf{I} \otimes \mathbf{F})\mathrm{vec}(\mathbf{A}'), \tag{2.10b}$$

where $\otimes$ denotes a Kronecker product.

The algorithm consists of two main steps. In the first step, we update $\mathbf{W}$ for fixed $\mathbf{A}'$. (To compute an initial estimate of $\mathbf{W}$, $\mathbf{A}'$ is initialized with arbitrary values.) Let $\mathbf{\Omega} = \mathbf{A} \otimes \mathbf{Z}^{(2)}$ and $\mathbf{w} = \mathrm{vec}(\mathbf{W})$ from (2.10a). There could be quite a few zero elements in $\mathbf{w}$, depending on the model to be fitted. Let $\mathbf{w}^*$ denote the vector formed by

eliminating zero elements from $\mathbf{w}$. Let $\mathbf{\Omega}^*$ denote the matrix formed by eliminating the columns of $\mathbf{\Omega}$ corresponding to the zero elements in $\mathbf{w}$. Then, we obtain the least squares estimate of $\mathbf{w}^*$ by

$$\hat{\mathbf{w}}^* = (\mathbf{\Omega}^{*\prime}\mathbf{\Omega}^*)^{-1}\mathbf{\Omega}^{*\prime}\text{vec}(\mathbf{Z}^{(1)}), \tag{2.11}$$

assuming that $\mathbf{\Omega}^{*\prime}\mathbf{\Omega}^*$ is nonsingular. (The regular inverse may be replaced by the Moore-Penrose inverse if $\mathbf{\Omega}^{*\prime}\mathbf{\Omega}^*$ is singular.) We can simply reconstruct the updated $\mathbf{w}$ from $\hat{\mathbf{w}}^*$ by putting back the zero elements to their original positions, and then the updated $\mathbf{W}$ from $\mathbf{w}$. We then obtain $\mathbf{F} = \mathbf{Z}^{(2)}\mathbf{W}$ and normalize it so that $\text{diag}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$. When more than one component are extracted from same sets of exogenous variables, columns of $\mathbf{F}$ are orthonomalized by the Gram-Schmidt orthonormalization method: the orthonormalized $\mathbf{F}$ is obtained by $\mathbf{F}\mathbf{R}_F^{-1}$, where $\mathbf{R}_F$ is obtained from the Cholesky factorization of $\mathbf{F}'\mathbf{F} = \mathbf{R}_F\mathbf{R}_F'$.

In the next step, $\mathbf{A}'$ is updated for fixed $\mathbf{W}$. Let $\mathbf{\Gamma} = \mathbf{I} \otimes \mathbf{F}$ and $\mathbf{a} = \text{vec}(\mathbf{A}')$ from (2.10b). We define $\mathbf{a}^*$ and $\mathbf{\Gamma}^*$ in a way similar to $\mathbf{w}^*$ and $\mathbf{\Omega}^*$ were obtained in the first step. For given $\mathbf{F}$, we obtain the least squares estimate of $\mathbf{a}^*$ by

$$\hat{\mathbf{a}}^* = (\mathbf{\Gamma}^{*\prime}\mathbf{\Gamma}^*)^{-1}\mathbf{\Gamma}^{*\prime}\text{vec}(\mathbf{Z}^{(1)}), \tag{2.12}$$

assuming that $\mathbf{\Gamma}^{*\prime}\mathbf{\Gamma}^*$ is nonsingular. (Again the regular inverse may be replaced by the Moore-Penrose inverse if $\mathbf{\Gamma}^{*\prime}\mathbf{\Gamma}^*$ is singular.) We can also easily recover the updated $\mathbf{a}$ and the updated $\mathbf{A}'$ from $\hat{\mathbf{a}}^*$. The above two steps are alternated until convergence is reached, that is, until the decrease in the function value falls below a certain threshold value, say 10^{-4}.

A few remarks concerning the proposed algorithm are in order. First of all, criterion (2.8) or equivalently (2.9) is consistently minimized by this algorithm. However, the algorithm does not guarantee that the obtained minimum is a global minimum. This so-called convergence to non-global minimum problem may be avoided in two ways (e.g., ten Berge, 1993). When we choose good (or rational) initial values, the function value is likely to start near to the global minimum, and it is more likely to obtain the global minimum. Here, we compute the PCA solutions of $\mathbf{Z}^{(2)}$, and use the principal coordinates as the rational start for $\mathbf{W}$. Then, initial values of $\mathbf{A}'$ are simply obtained by the least squares estimate, given $\mathbf{W}$. The second possible remedy against the non-global minimum problem is to repeat the ALS procedure with many random initial starts. The obtained function values after convergence are compared, and the smallest one is chosen as the global minimum. We consider parameter estimates associated with the smallest function value as the optimal ones.

When n is large relative to q, the above algorithm may be made more efficient by the following procedure. Let $\mathbf{Z}^{(2)} = \mathbf{QR}'$ be portion of the QR decomposition of $\mathbf{Z}^{(2)}$, pertaining to the column space of $\mathbf{Z}^{(2)}$, where $\mathbf{Q}$ is an n by q orthonormal matrix, so that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$, and $\mathbf{R}'$ is a q by n upper-triangular matrix. Then, (2.8) can be rewritten as follows:

$$\begin{aligned} f &= \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{QR}'\mathbf{B}) \\ &= \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{QQ}'\mathbf{Z}^{(1)} + \mathbf{QQ}'\mathbf{Z}^{(1)} - \mathbf{QR}'\mathbf{B}) \\ &= \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{QQ}'\mathbf{Z}^{(1)}) + \mathrm{SS}(\mathbf{Q}(\mathbf{Q}'\mathbf{Z}^{(1)} - \mathbf{R}'\mathbf{B})), \end{aligned} \tag{2.13}$$

where $\mathbf{B} = \mathbf{WA}'$. The first term of the right-hand side in (2.13) does not depend on $\mathbf{B}$,

and minimizing (2.13) reduces to minimizing

$$f^* = \mathrm{SS}(\mathbf{Q}(\mathbf{Q}'\mathbf{Z}^{(1)} - \mathbf{R}'\mathbf{B})) = \mathrm{SS}(\mathbf{Q}'\mathbf{Z}^{(1)} - \mathbf{R}'\mathbf{B}). \quad (2.14)$$

It is more efficient to minimize (2.14) instead of (2.8) because the size of $\mathbf{R}'$ is usually much smaller than $\mathbf{Z}^{(2)}$. Moreover, this procedure allows us to use covariance matrices, or correlation matrices instead of data matrices, because $\mathbf{Z}^{(2)\prime}\mathbf{Z}^{(2)} = \mathbf{R}\mathbf{R}'$ and $\mathbf{Z}^{(2)\prime}\mathbf{Z}^{(1)} = \mathbf{R}\mathbf{Q}'\mathbf{Z}^{(1)}$, so that $\mathbf{Q}'\mathbf{Z}^{(1)}$ can be obtained by $\mathbf{R}^{-1}\mathbf{Z}^{(2)\prime}\mathbf{Z}^{(1)}$. This is often beneficial because in the published literature data sets are frequently provided in the form of covariance or correlation matrices. When we deal with the covariance or correlation matrices, however, we could not obtain $\mathbf{F}$ (Takane, Kiers, & de Leeuw, 1995).

It may also be worthwhile to compare the estimation procedure with that of covariance structure analysis and PLS. Covariance structure analysis is typically based on maximum likelihood estimation (MLE) although least squares estimation (GLS or ULS) is also used. The MLE estimates parameters in such a way that they maximize the probability of obtaining the observed data (or the sample covariance matrix derived from the data) given the hypothesized model (reflected in the implied covariances) under certain distributional assumptions. In contrast, our method estimates model parameters by minimizing the sum of squares of discrepancies between the observed endogenous variables and their predicted counterparts from the exogenous variables without any explicit distributional assumptions. This is similar to PLS. However, PLS does not optimize a global fitting criterion; instead, it repeats solving a series of simple

or multiple regression analysis problems. Our method, on the other hand, solves a global optimization problem by consistently minimizing (2.8) or (2.9). This is akin to covariance structure analysis that maximizes a likelihood function. Due to the absence of a global minimization criterion, PLS defines convergence as a sort of equilibrium (i.e., the point at which no difference between the previous and current estimates), whereas our method defines convergence as the decrease in the function value falling below a certain threshold.

Many fit indices used in covariance structure analysis (e.g., χ^2, GFI, CFI, RMSEA, AIC, etc.) can not be used to assess the goodness of fit of a model in our method, since those indices are valid only under specific distributional assumptions such as multivariate normality. Instead, in our 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, or equivalently, by the total variance of the endogenous variables minus its unexplained variance. This is given by

$$\text{Fit} = 1 - \frac{\text{SS}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')}{\text{SS}(\mathbf{Z}^{(1)})}. \qquad (2.15)$$

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.

We may use resampling methods such as the jackknife and the bootstrap methods to calculate standard errors of parameter estimates. In our method, the standard errors are estimated by the bootstrap method (Efron, 1982; Efron & Tibshirani, 1998). The bootstrapped standard errors can be used to assess the reliability of the parameter

estimates. The critical ratios (i.e., the parameter estimates divided by their standard errors) can be used to test 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 .05 level).

2.4. Imposing Additional Constraints on Parameters

We may be interested in testing various structural hypotheses regarding **W** and/or $\mathbf{A}'$. For instance, we may examine the hypothesis that some elements in **W** are equal or that some elements in $\mathbf{A}'$ are equal, and so on. A variety of other structural hypotheses regarding **W** and $\mathbf{A}'$ can be incorporated in the form of linear constraints (Böckenholt & Böckenholt, 1990; Böckenholt & Takane, 1994; Takane & Shibayama, 1991; Takane, Yanai, & Mayekawa, 1991; ter Braak, 1986; Yanai, 1986, etc.).

The linear constraints may be specified by either the reparametrization or the null-space method (Böckenholt & Takane, 1994; Takane, Yanai, & Mayekawa, 1991). The former method specifies the space spanned by column vectors of a constraint matrix, while the latter specifies its ortho-complement space. In our method, all linear constraints are imposed by the reparametrization method. Let **H** denote a matrix of linear constraints on **a**. In the final step of the ALS algorithm, we incorporate **H** into **a** as follows:

$$\mathbf{a} = \mathbf{H}\boldsymbol{\alpha}, \tag{2.16}$$

for some $\boldsymbol{\alpha}$. An LS estimate of $\boldsymbol{\alpha}$ is then given by

$$\hat{\boldsymbol{\alpha}} = (\mathbf{H}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{H})^{-1}\mathbf{H}'\boldsymbol{\Gamma}'\text{vec}(\mathbf{Z}^{(1)}), \tag{2.17}$$

which leads to

$$\hat{\mathbf{a}} = \mathbf{H}\hat{\boldsymbol{\alpha}} = \mathbf{H}(\mathbf{H}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{H})^{-1}\mathbf{H}\boldsymbol{\Gamma}'\text{vec}(\mathbf{Z}^{(1)}). \tag{2.18}$$

This approach is called the projection method (see Seber, 1984, pp. 403-405; Takane, Yanai, & Mayekawa, 1991).

It is sometimes easier to specify constraints in the null-space form (e.g., equality or zero constraints). In such cases, the constraints are first expressed in the null-space form, and then transformed into the reparametrization form. The transformation is straightforward. Let

$$\mathbf{P}'\mathbf{a} = \mathbf{0} \tag{2.19}$$

represent the constraints in the null space form. Suppose that the first and the last elements of $\mathbf{a}$ are equal, then, $\mathbf{P}'$ comes down to a vector whose first element is 1, the last element is -1, and the other elements are zeros. We may reparametrize (2.19) into the form of (2.16) by defining $\mathbf{H} = \mathbf{I} - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-}\mathbf{P}'$. This implies that $\text{Ker}(\mathbf{P}') = \text{Sp}(\mathbf{H})$, where $\text{Ker}(\mathbf{P}')$ denotes the null space of $\mathbf{P}'$, and $\text{Sp}(\mathbf{H})$ denotes the space spanned by the column vectors of $\mathbf{H}$. Linear constraints can be imposed on $\mathbf{w}$ in a similar way. The validity of certain hypotheses may be empirically investigated by comparing fits of the constrained and unconstrained solutions. The standard errors and critical ratios of the obtained parameter estimates can also be used to evaluate the hypotheses.

In the special case in which the constraints on $\mathbf{A}'$ ($\mathbf{a} = \mathbf{H}\boldsymbol{\alpha}$) can be written as

$\mathbf{A}^{*\prime}\mathbf{H}$, the extended redundancy analysis model may be expressed as

$$\begin{aligned}\mathbf{Z}^{(1)} &= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}^{*\prime}\mathbf{H} + \mathbf{E} \\ &= \mathbf{Z}^{(2)}\mathbf{B}^{*}\mathbf{H} + \mathbf{E},\end{aligned} \tag{2.20}$$

where $\mathbf{B}^{*} = \mathbf{W}\mathbf{A}^{*\prime}$. This is essentially the same as generalized multivariate ANOVA (GMANOVA) or the growth curve models (Potthoff & Roy, 1964). If $\mathbf{B}^{*}$ is further assumed to have reduced rank, i.e., rank($\mathbf{B}^{*}$) < min(p,q), (2.20) comes down to reduced-rank growth curve models proposed by Reinsel and Velu (1998).

Furthermore, let $\mathbf{Z}^{(2)} = [\mathbf{Z}_1, \cdots, \mathbf{Z}_I], \mathbf{W} = [\mathbf{W}_1, \cdots, \mathbf{W}_I]$, $\mathbf{A}^{*\prime} = [\mathbf{A}_1^{*\prime}, \cdots, \mathbf{A}_I^{*\prime}]$, $\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_I]$, Then, (2.20) may be rewritten as

$$\mathbf{Z}^{(1)} = \sum_{i=1}^{I} \mathbf{Z}_i \mathbf{B}_i^{*} \mathbf{H}_i + \mathbf{E}, \tag{2.21}$$

where $\mathbf{B}_i^{*} = \mathbf{W}_i \mathbf{A}_i^{*\prime}$. This is analogous to the constrained component analysis model proposed by Takane, Kiers, and de Leeuw (1995), where different sets of constraints are imposed on different dimensions (DCDD) of the data matrix. This indicates that the constrained extended redundancy analysis reduces to the DCDD type-constrained component analysis, as the number of subsets (or submatrices) of $\mathbf{Z}^{(2)}$ and $\mathbf{H}$ are identical. Hence, our method may be viewed as a generalization of the constrained component analysis model.

2.5. Example: The Basic Health Indicator Data

In this section, we present an example to demonstrate the feasibility of the proposed method. The example is part of the so-called basic health indicator data

collected by the World Health Organization in the United Nations. They appear in the 1999 World Health Report and also are available through the internet (http://www.who.int). From the entire data set, we only used six variables measured in different countries. The six observed variables were as follows: (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 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 was 51, which corresponded with the number of countries for which the data were available.

We assumed two latent variables 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. The specified two latent variable model is depicted in Figure 2.4. For this model, $\mathbf{W}$ and $\mathbf{A}'$ were identical to those in (2.4). Using extended redundancy analysis, this model was fitted to the data. Results are provided in Figure 2.5. The bootstrapped standard errors and the critical ratios of parameter estimates obtained with 100 boostrap samples are given in Table 2.1.

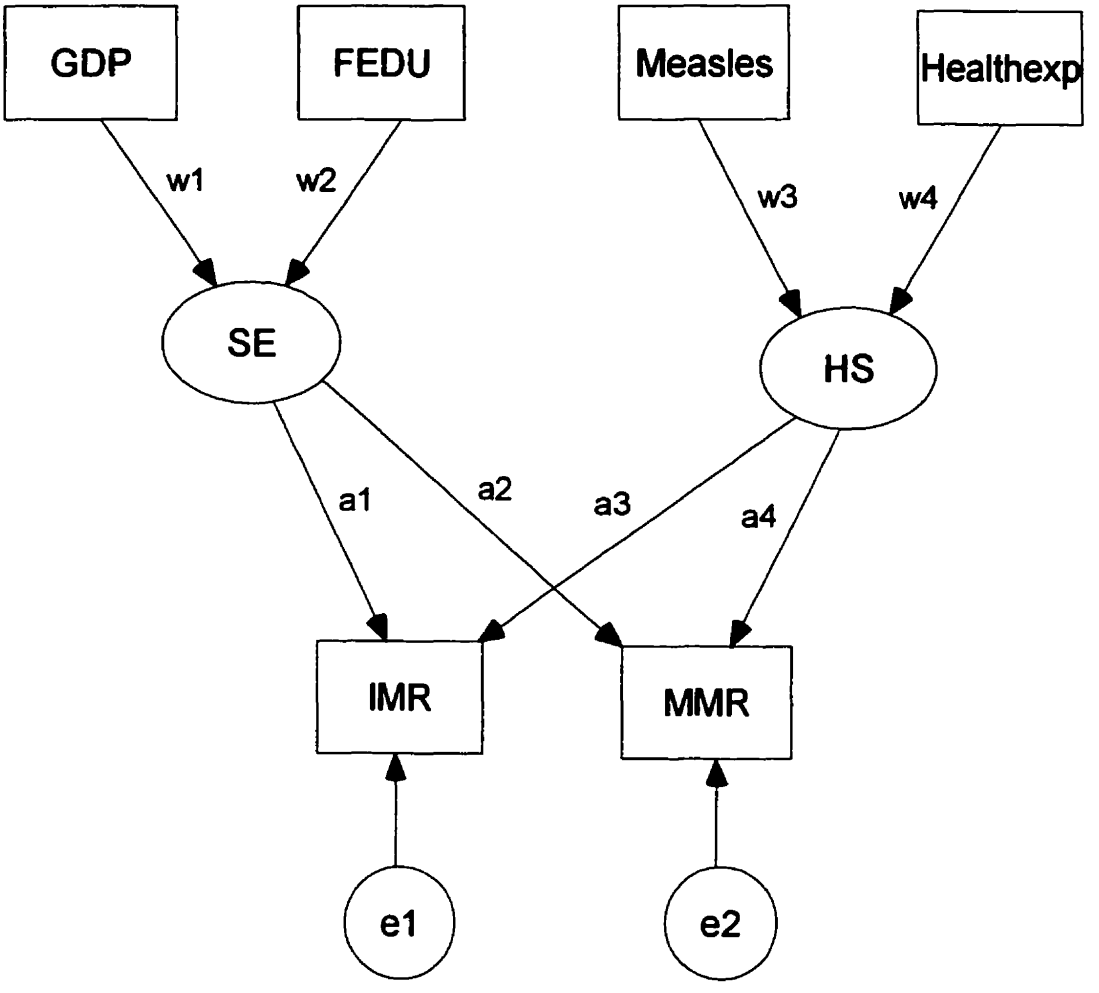

Figure 2.4. The two latent variable model for the WHO data.

Fit = .6512

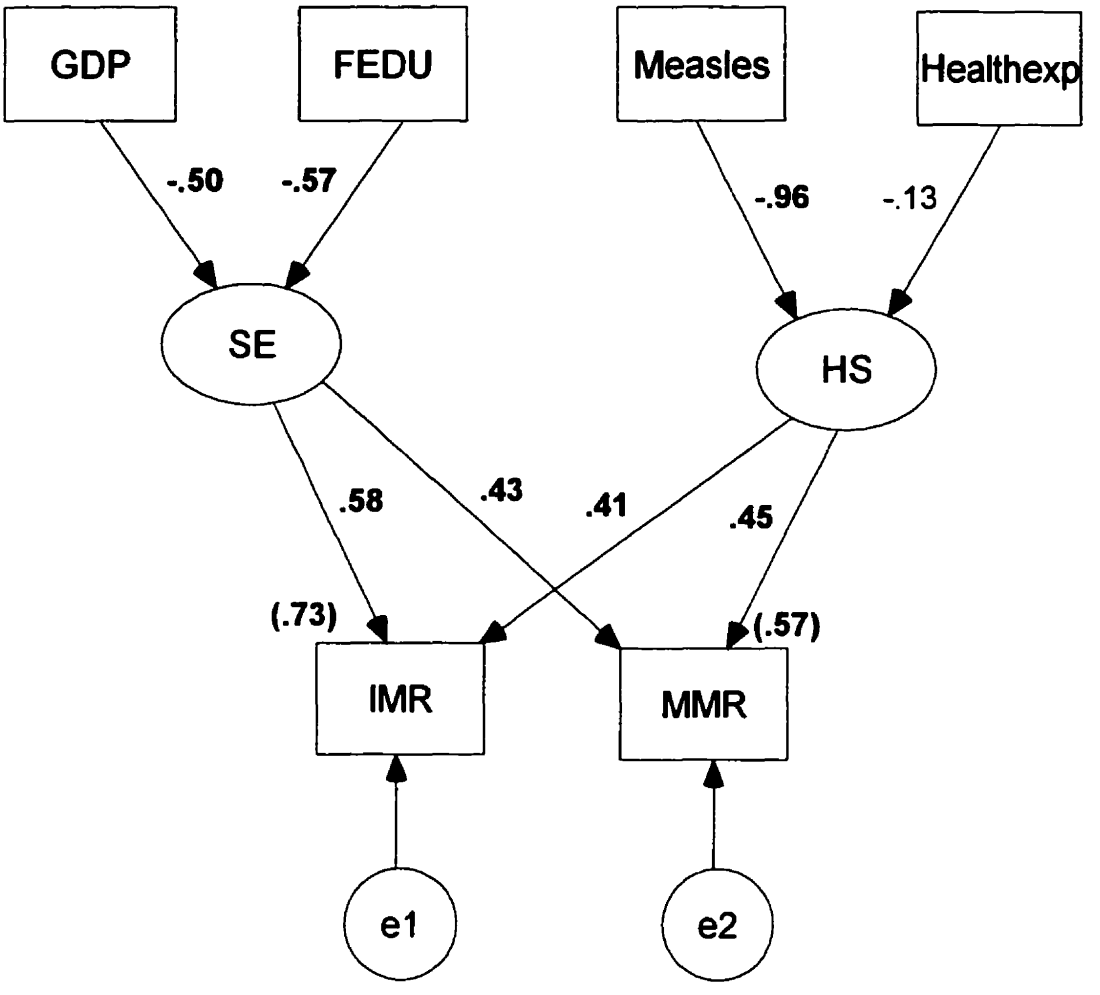

Figure 2.5. The two latent variable model for the WHO data (output).

Table 2.1. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the two latent variable model for the WHO data.

	Estimate	SE	CR
w_1	-.50	.17	-2.9
w_2	-.57	.16	-3.6
w_3	-.96	.12	-8.0
w_4	-.13	.24	-0.5
a_1	.58	.10	5.8
a_2	.43	.10	4.3
a_3	.41	.09	4.6
a_4	.45	.11	4.1

The goodness of fit of the model was equal to .6512, indicating that about 65% of the total 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 (14.5), 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. This indicated that about 73% of the variance of IMR and about 57% of the variance of MMR were explained by the two latent variables. They also turned out significant according to their bootstrapped critical ratios (18.7 and 10.0 for IMR and MMR, respectively).

In Figure 2.5 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 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 variable was .47. It turned out to be significant in terms of its bootstrapped critical ratio (3.9).

Given the solutions in the two latent variable model, we further assumed that the component weight for Healthexp was equal to zero (i.e., $w_4 = 0$). This additional

assumption could be incorporated by imposing zero constraints on **w**, as explained in Section 2.4. Let $\mathbf{p}'$ denote the vector of the zero constraints. The $\mathbf{p}'$ was defined as

$$\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}.$$

It was required that $\mathbf{p}'\mathbf{w} = 0$. Then, the $\mathbf{p}'$ was transformed into a reparametrization form by the procedure described in Section 2.4. Results of fitting the constrained model are presented in Figure 2.6. The bootstrapped standard errors and the critical ratios obtained with 100 bootstrap samples are given in Table 2.2. Due to the zero constraint, the component weight for Healthexp was shown as zero in Figure 2.6. This indicated that HS was solely defined in terms of Measles, which further implied that HS was completely equivalent to Measles. It thus was analogous to eliminating HS from the analysis, and hypothesizing direct effects of Measles on the endogenous variables. Measles was found to have negative direct effects on IMR (-.40) and MMR (-.43). The fit of the constrained model was .6491. This was almost the same as that of the unconstrained model. It turned out to be significant in terms of its bootstrapped critical ratio (13.5). The squared multiple correlations for IMR and MMR were equal to .73 and .57, respectively, which were essentially the same as those obtained from the unconstrained case. They turned out significant in terms of their bootstrapped critical ratios (18.5 and 8.9 for IMR and MMR, respectively). The correlation between the latent variables became somewhat smaller (.41) than that from the unconstrained case (.47). This is due to the elimination of Healthexp from HS. Yet the correlation turned out to be significant (the bootstrapped critical ratio = 3.2). Despite the zero constraint,

Fit = .6491

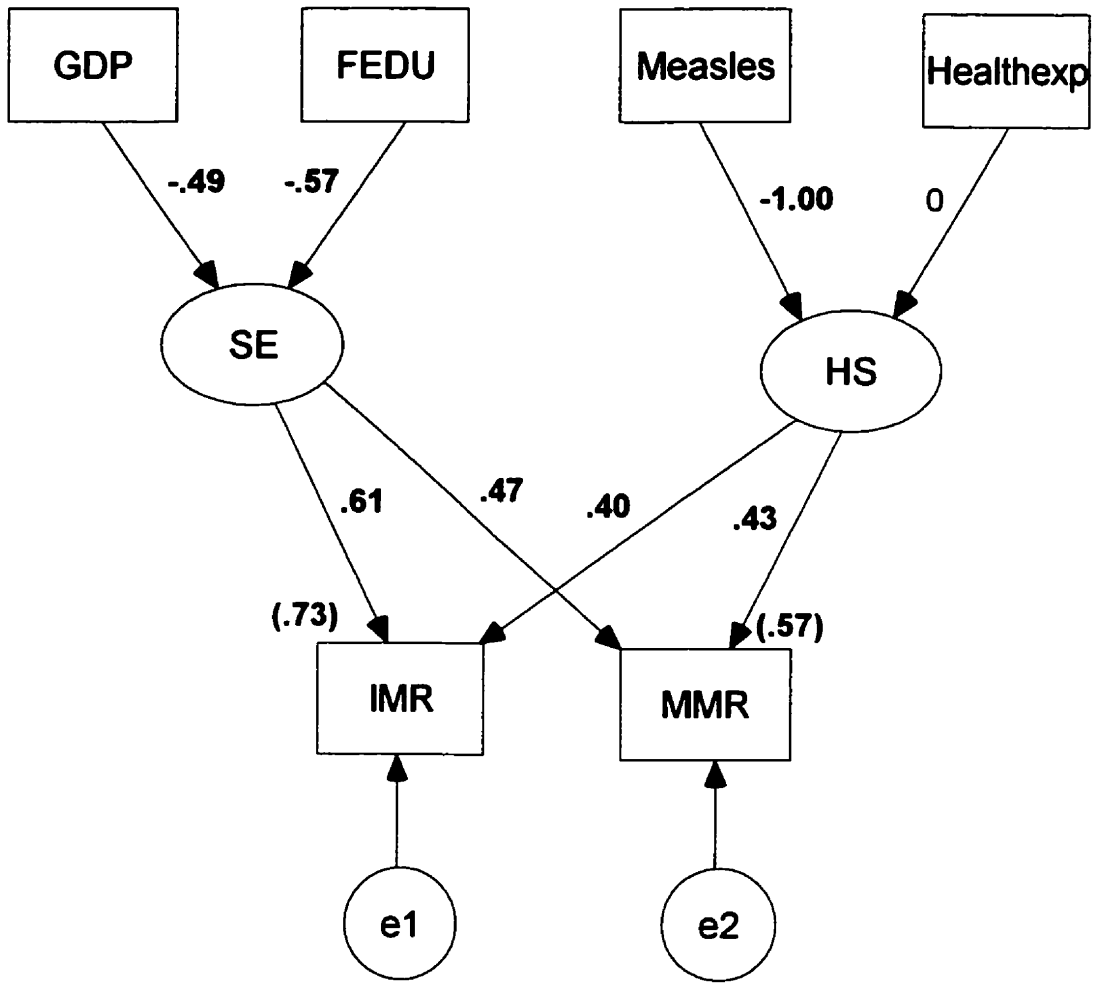

Figure 2.6. The constrained two latent variable model for the WHO data (output).

Table 2.2. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the constrained two latent variable model for the WHO data.

	Estimate	SE	CR
w_1	-.49	.16	-3.1
w_2	-.57	.16	-3.6
a_1	.61	.11	5.5
a_2	.47	.13	3.6
a_3	.40	.11	3.6
a_4	.43	.15	2.9

however, interpretations of the solutions from the constrained model seemed to be essentially the same as those from the unconstrained model. Therefore, it may be safe to say that the additional assumption on the component weight for Healthexp is acceptable. This allows for simpler interpretations of the obtained solutions by reducing the number of parameters.

Chapter 3

Some Extensions

3. 1. Three Possible Extensions

The ERA model in (2.6) can be readily extended in various ways. In particular, we discuss ways to handle higher-order latent variables (i.e., latent variables nested within other latent variables), direct effects of observed variables on other observed variables, and multi-sample comparisons. It is also shown that the extended models can be expressed in essentially the same form as (2.1), and essentially the same estimation procedure can be applied to fit them.

To include the Kth-order latent variables, the ERA model may be expressed as follows.

$$\mathbf{Z}^{(1)} = \mathbf{Z}^{(2)}\tilde{\mathbf{W}}\mathbf{A}^{'} + \mathbf{E}, \tag{3.1}$$

where

$$\tilde{\mathbf{W}} = \prod_{k=1}^{K} \mathbf{W}^{(k)}. \tag{3.2}$$

In (3.2), $\mathbf{W}^{(k)}$ denotes the matrix of component weights for the kth-order latent variables ($k = 1, \cdots, K$), and each kth-order latent variable is restricted to be of unit

length for identification. Model (3.1) is essentially the same as (2.6). Hence, a similar ALS algorithm can be used to fit the model. In this case, however, a set of parameter matrices are split into $K+1$ matrices, that is, $\mathbf{W}^{(k)}$'s and $\mathbf{A}'$, and we update them alternately until convergence is obtained. For instance, an ERA model with second-order latent variables may be written as

$$\begin{aligned} \mathbf{Z}^{(1)} &= \mathbf{Z}^{(2)}\mathbf{W}^{(1)}\mathbf{W}^{(2)}\mathbf{A}' + \mathbf{E}, \\ &= \mathbf{Z}^{(2)}\tilde{\mathbf{W}}\mathbf{A}' + \mathbf{E}, \end{aligned} \tag{3.3}$$

where $\tilde{\mathbf{W}} = \mathbf{W}^{(1)}\mathbf{W}^{(2)}$. To estimate $\tilde{\mathbf{W}}$ and $\mathbf{A}'$, in (3.3), we may update $\mathbf{W}^{(1)}$ for fixed $\mathbf{A}'$ and $\mathbf{W}^{(2)}$, normalize $\mathbf{Z}^{(2)}\mathbf{W}^{(1)}$, update $\mathbf{W}^{(2)}$ for fixed $\mathbf{A}'$ and $\mathbf{W}^{(1)}$, normalize $\mathbf{Z}^{(2)}\mathbf{W}$, and update $\mathbf{A}'$ for fixed $\mathbf{Z}^{(2)}\mathbf{W}$ in each iteration. An example of a model with higher-order latent variables will be given in the next section.

The ERA model including the direct effects may also be written in the same form as (2.6). For example, we suppose that in (2.4), $\mathbf{z}_1$ of $\mathbf{Z}^{(2)}$ has a direct effect on $\mathbf{z}_5$ of $\mathbf{Z}^{(1)}$ in addition to the effect on $\mathbf{z}_5$ through $\mathbf{f}_1$. To include this effect, we may write the ERA model as follows.

$$\mathbf{Z}^{(1)} = \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E}, \tag{3.4}$$

where

$$\mathbf{W} = \begin{bmatrix} 1 & w_{11} & 0 \\ 0 & w_{12} & 0 \\ 0 & 0 & w_{21} \\ 0 & 0 & w_{22} \end{bmatrix},$$

and

$$\mathbf{A}' = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

Model (3.4) is a simple variant of (2.6), where the only distinction is in that $\mathbf{W}$ and $\mathbf{A}'$ contain an extra (first) column and row, respectively, in order to represent a direct effect of $\mathbf{z}_1$ on $\mathbf{z}_5$. In this way, we can readily incorporate direct effects of any observed variables in $\mathbf{Z}^{(2)}$. The supplementary column of $\mathbf{W}$ consists of all fixed elements, zeros or unities, whereas the corresponding row of $\mathbf{A}'$ has free parameters to be estimated as well as fixed elements. Essentially the same estimation procedure can be applied to fit (3.4). To update $\mathbf{W}$, however, the effect corresponding to the unit elements in $\mathbf{W}$ is subtracted from $\mathbf{Z}^{(1)}$, the unit elements in $\mathbf{W}$ are eliminated from vec($\mathbf{W}$) like the zero elements, and the corresponding columns in $\boldsymbol{\Omega}$ are eliminated. Then $\mathbf{w}^*$ is obtained by equation (2.11). The unit elements are then refilled when $\mathbf{W}$ is formed from $\mathbf{w}^*$. To demonstrate the addition of direct effects to the ERA model, an example will be presented in Section 3.2.

We may also be interested in fitting a single ERA model to more than one sample simultaneously. Such a simultaneous analysis enables us to test various hypotheses concerning the relationships among parameters across samples. Suppose that J samples are fitted by the same ERA model,

$$\begin{aligned}\mathbf{Z}_j^{(1)} &= \mathbf{Z}_j^{(2)}\mathbf{W}_j\mathbf{A}_j' + \mathbf{E}_j \\ &= \mathbf{F}_j\mathbf{A}_j' + \mathbf{E}_j, \end{aligned} \tag{3.5}$$

where $\mathbf{F}_j = \mathbf{Z}_j^{(2)}\mathbf{W}_j$ ($j = 1, \cdots, J$) with $\text{diag}(\mathbf{F}_j'\mathbf{F}_j) = \mathbf{I}$. Here the structures of $\mathbf{W}_j$ and $\mathbf{A}_j'$ may be identical across J samples. Model (3.5) may be re-expressed as

$$\begin{bmatrix} \mathbf{Z}_1^{(1)} \\ \vdots \\ \mathbf{Z}_J^{(1)} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1^{(2)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_J^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{W}_J \end{bmatrix} \begin{bmatrix} \mathbf{A}_1' \\ \vdots \\ \mathbf{A}_J' \end{bmatrix} + \begin{bmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_J \end{bmatrix}. \tag{3.6}$$

It can thus be expressed as a single equation,

$$\dot{\mathbf{Z}}^{(1)} = \dot{\mathbf{Z}}^{(2)}\dot{\mathbf{W}}\mathring{\mathbf{A}}' + \dot{\mathbf{E}}, \tag{3.7}$$

where $\dot{\mathbf{Z}}^{(1)} = [\mathbf{Z}_1^{(1)}, \cdots, \mathbf{Z}_J^{(1)}]'$, $\dot{\mathbf{Z}}^{(2)} = \text{diag}[\mathbf{Z}_1^{(2)}, \cdots, \mathbf{Z}_J^{(2)}]$, $\dot{\mathbf{W}} = \text{diag}[\mathbf{W}_1, \cdots, \mathbf{W}_J]$, $\mathring{\mathbf{A}}' = [\mathbf{A}_1', \cdots, \mathbf{A}_J']'$, and $\dot{\mathbf{E}} = [\mathbf{E}_1, \cdots, \mathbf{E}_J]'$. Model (3.7) is essentially the same as (2.6), and the same optimization procedure can be used. To test structural hypotheses concerning the parameters across J samples (e.g., equality among some parameters across sample), J sets of parameters in $\mathbf{W}_j$ and $\mathbf{A}_j'$ can be regarded as a single set of parameters in $\dot{\mathbf{W}}$ and $\mathring{\mathbf{A}}'$ as in a single sample, so that the same procedure in Section 2.4 can be used. More specifically, we update the vector, say $\dot{\mathbf{w}}^*$, which is formed by eliminating zero elements from $\text{vec}(\dot{\mathbf{W}})$, by equation (2.11), and recover the updated $\dot{\mathbf{W}}$. We normalize $\mathbf{F}_j$ for each sample, such that $\text{diag}(\mathbf{F}_j'\mathbf{F}_j) = \mathbf{I}$. Then $\mathring{\mathbf{A}}'$ is updated from $\mathring{\mathbf{a}}^*$ formed by eliminating zero elements from $\text{vec}(\mathring{\mathbf{A}}')$. We may easily compare the

means of the components, $\mathbf{F}_j$, across samples. In this case, however, the unstandardized data should be analyzed instead of the standardized data since the means are *a priori* eliminated in the standardized data. The set of exogenous variables should also include a constant term. An example of the simultaneous analysis of several samples will be provided in the next section.

3. 2. Examples

The basic health indicator data are again used to demonstrate the extensions described in the previous section. We specify two models more complicated than the two latent variable model shown in Section 2.5: One adds a second-order latent variable and the other a direct effect. Another data set, called social function data, is employed to show the feasibility of simultaneous analysis of several samples or multi-sample comparisons. Note that the two structural equation models for the basic health indicator data are specified without well-grounded theories or hypotheses to support our model specification. They are simply improvised to illustrate the usefulness of the proposed method. On the other hand, an empirically well-motivated structural equation model is fitted in the analysis of the second data set.

3.2.1. The Basic Health Indicator Data

To demonstrate the feasibility of incorporating higher-order latent variables, we supposed a second-order latent variable nested within the two first-order latent variables in the two latent variable model shown in Figure 2.4. The second-order latent variable was named 'combined effect (CE)'. The specified second-order latent variable

model is provided in Figure 3.1. In the model, $\mathbf{W}^{(1)}$ was analogous to $\mathbf{W}$ in (2.4), and $\mathbf{W}^{(2)}$ and $\mathbf{A}'$ were specified as follows.

$$\mathbf{W}^{(2)} = \begin{bmatrix} w_5 \\ w_6 \end{bmatrix},$$

and

$$\mathbf{A}' = \begin{bmatrix} a_1 & a_2 \end{bmatrix}.$$

Results of fitting the model is presented in Figure 3.2. The standard errors and critical ratios of the parameter estimates obtained from the bootstrap method with 100 bootstrap samples are provided in Table 3.1. This model showed a slightly worse fit (.6492) than that without the second-order latent variable. Yet it turned out to be significant in terms of its bootstrapped critical ratio (15.0). Both component weights of SE and HS (.67 and .50, respectively) for CE were found to be significant and positive. The loadings of CE on IMR (.87) and MMR (.77) turned out to be significant and positive as well. This indicates that CE is a negatively combined effect for health supports, so that a large value on CE is likely to increase the possibility of infant and maternal deaths. The squared multiple correlations of IMR (.73) and MMR (.57) turned out to be significant in terms of their bootstrapped critical ratios, 21.9 and 9.9, respectively. The correlation between the latent variables was .46. It also turned out to be significant (the bootstrapped critical ratio = 3.8).

To illustrate the effect of adding direct effects, we assumed that GDP had a direct

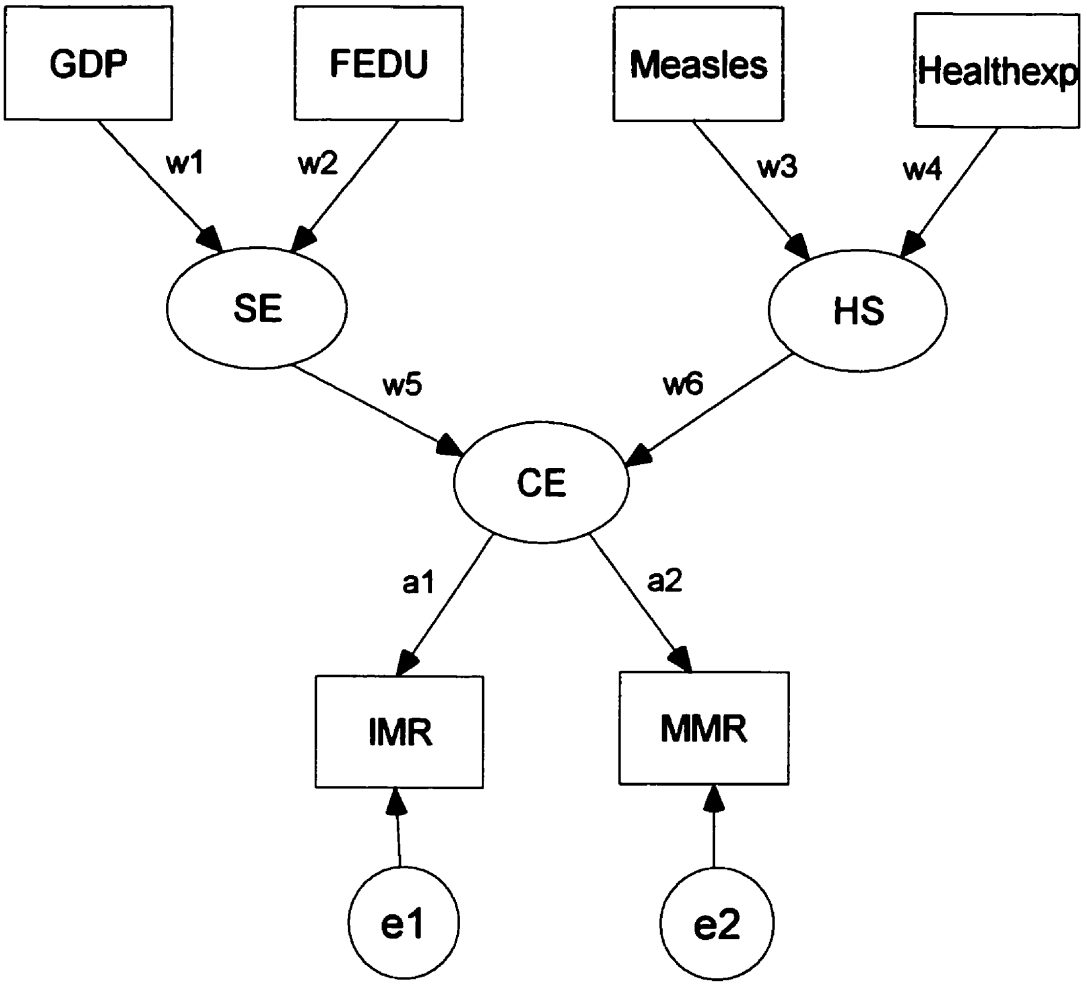

Figure 3.1. The second-order latent variable model for the WHO data.

Fit = .6492

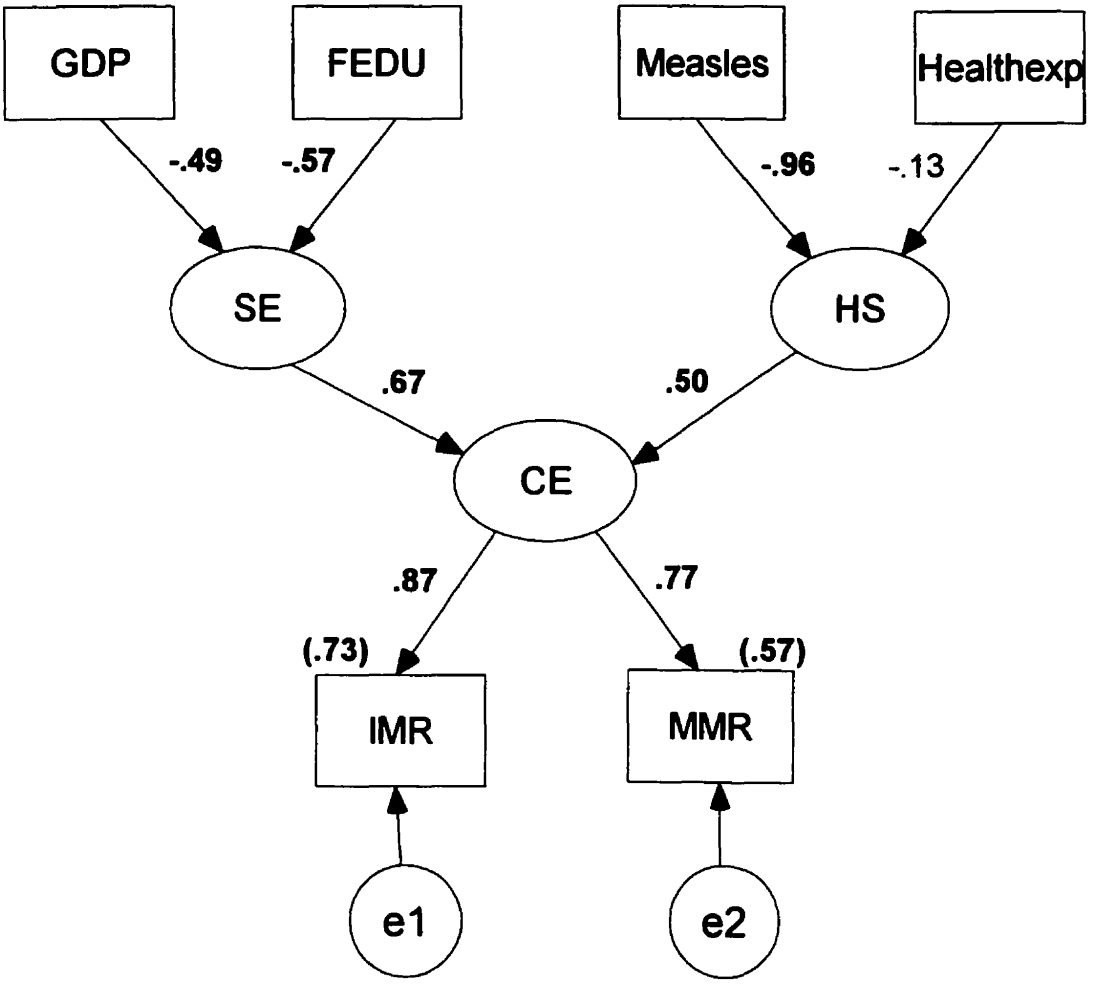

Figure 3.2. The second-order latent variable model for the WHO data (output).

Table 3.1. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the second-order latent variable model for the WHO data.

	Estimates	SE	CR
w_1	-.49	.18	-2.7
w_2	-.57	.18	-3.2
w_3	-.96	.13	-7.4
w_4	-.13	.23	-0.6
w_5	.67	.13	5.2
w_6	.50	.13	3.8
a_1	.87	.02	43.5
a_2	.77	.04	19.3

effect on IMR in addition to its effect through SE. The two latent variable model with the direct effect is depicted in Figure 3.3, in which the structures of $\mathbf{W}$ and $\mathbf{A}'$ were equivalent to those in (2.20). Results of fitting the model is presented in Figure 3.4. The standard errors and critical ratios of the parameter estimates obtained from the bootstrap method with 100 bootstrap samples are presented in Table 3.2. The model showed almost the same fit (.6513) as that without the direct effect, which turned out to be significant in terms of its critical ratio (15.1). This seems to be compatible with that the direct effect of GDP was non-significant. The effect of SE on IMR turned out to be nonsignificant as well. It is due to the effect of adding the direct of GDP on IMR to the model. The squared multiple correlations of IMR and MMR were .73 and .57, respectively. The squared multiple correlation of IMR is quite similar to that obtained from the model without the direct effect. This is also consistent with the nonsignificant direct effect of GDP on IMR. Both squared multiple correlations turned out to be significant in terms of their critical ratios (17.1 and 11.6 for IMR and MMR, respectively). The correlation between the latent variables (.47) turned out to be significant as well (the bootstrapped critical ratio = 3.4).

3.2.2. The Social Function Data

The second example was obtained from Park's (1996) social function data. Park (1996) studied cultural differences in the functional and structural aspects of social functions between South Korean and German adolescents. She assessed seven social functions, such as attachment, self-validation, intimacy, guidance, control, conflict, and comparison, that seemed to play important roles in the development of adolescents.

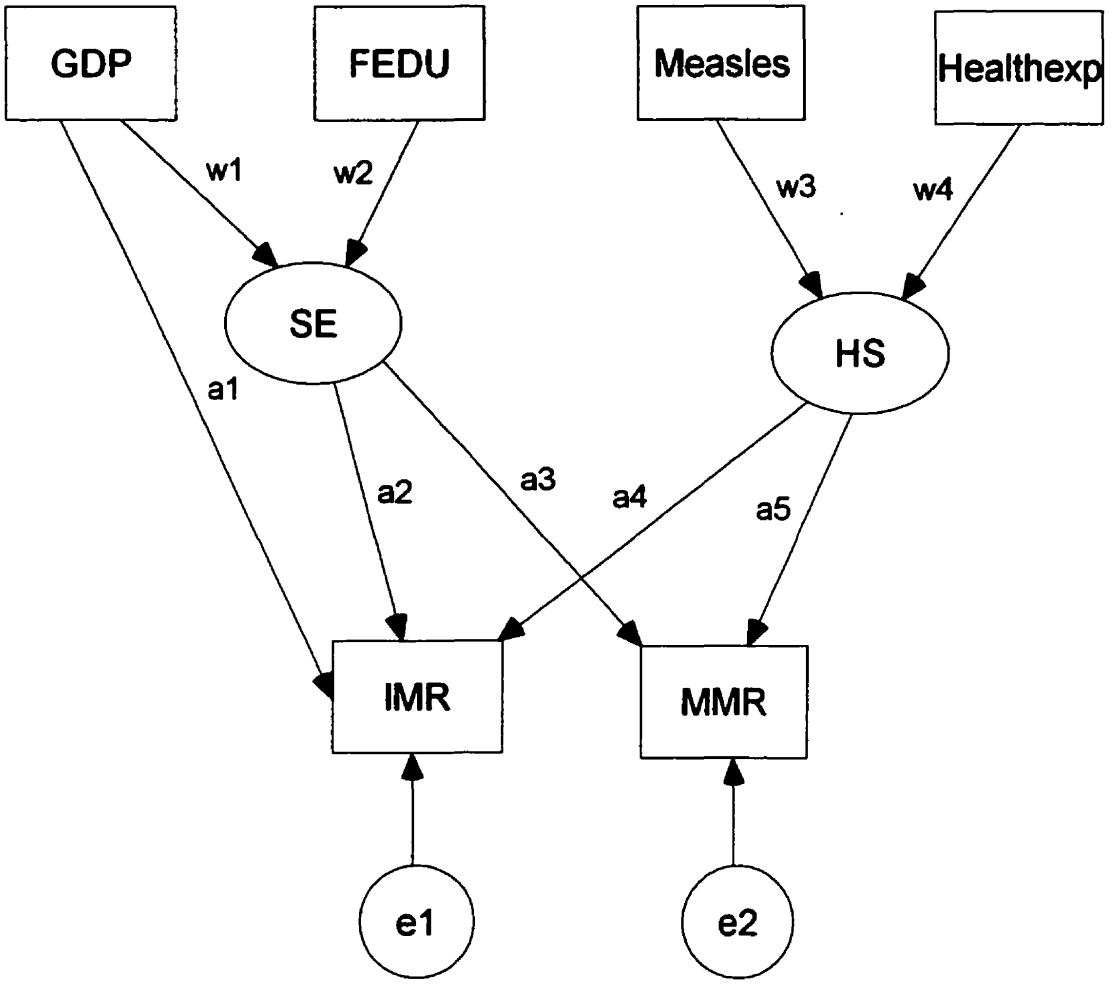

Figure 3.3. The two latent variable model for the WHO data with a direct effect added.

Fit = .6513

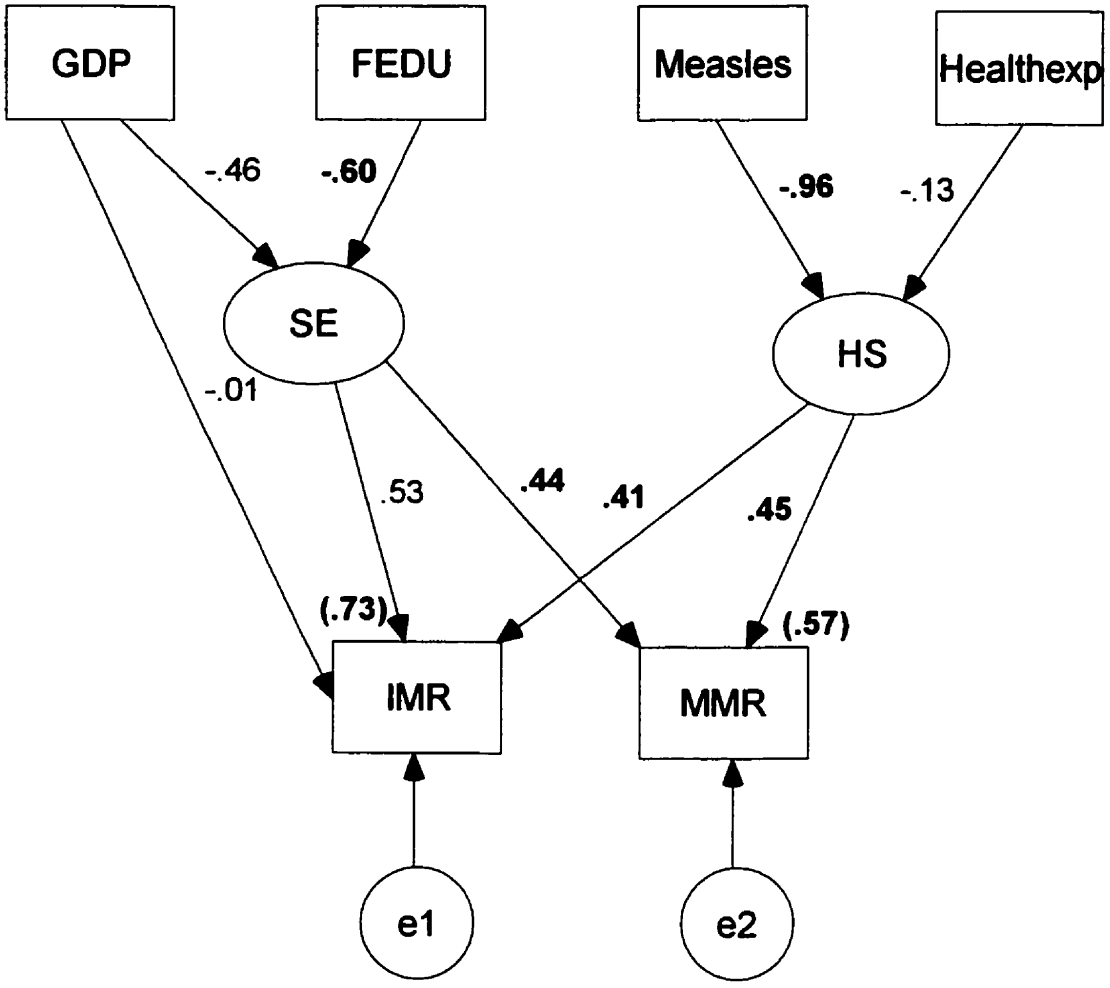

Figure 3.4. The two latent variable model for the WHO data with a direct effect added (output).

Table 3.2. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the two latent variable model for the WHO data with a direct effect added.

	Estimates	SE	CR
w_1	-.46	.25	-1.8
w_2	-.60	.24	-2.5
w_3	-.96	.08	-12.0
w_4	-.13	.21	-0.6
a_1	-.01	.21	-0.0
a_2	.53	1.46	0.4
a_3	.44	.10	4.4
a_4	.41	.11	3.7
a_5	.45	.13	3.5

(See Park (1996) for detailed explanations about those social functions.)

Fifty-nine South Korean adolescents and sixty German adolescents participated in her study. The South Korean participants were from a junior high school in Seoul (29 males and 30 females) and the German participants from 3 Gymnasiums in Kohn (30 males and 30 females) . The mean age of the South Korean participants was 14 years and 1 month old and that of the German participants was 14 years and 2 months old. A self-report questionnaire was administered to measure the degree of the social functions to which South Korean and German adolescents would be exposed. The self-report questionnaire consisted of 18 items on adolescents' behaviors related to the seven social functions specified above. Most of the items were constructed on the basis of analysis of diaries of 10 South Korean adolescents who were 11 years old. Items indicating each social function were as follows: (1) attachment: "Did you feel happy when you did something with him/her together?", "Did you feel happy when he/she gave you something , or you could give him/her something?", "Did you worry about him/her?", (2) self-validation: "Did you gain recognition from him/her?", "Did you gain praise from him/her?", (3) intimacy: "Did you talk to him/her about what happened to you today?", "Did you speak to him/her about your innermost feelings?", (4) guidance: "Did he/she help you solve problems?", "Did the person advise you?", "Did you speak to the person about your problem?", (5) control: "Did the person say to you that you must do something (e.g., studying, making up the room, etc.)?", "Did the person punish you?", (6) conflict: "Did you quarrel with the person?", "Did you hate the person?", (7) comparison: "Did you feel superiority to the person?", "Did you feel inferiority to the

person", "Were you envious of the person?" To assess a dependent measure, 'trust', a corresponding item was also included in the questionnaire: "How often will the person help you if you are in trouble?" .

The South Korean adolescent sample was collected in the fall semester of 1992 and the German adolescent sample was in the spring and fall semester of 1993 and 1994, respectively. All participants were first asked to make a list of significant persons whom they have contacted at least once per week by various means, including by phone and by mail. The significant persons on the list were considered as social network members. The participants were then asked to answer each item of the questionnaire for each person on the list. The questionnaire was filled out every day in the evening, and was collected daily in school for a week. About 50% of the participants could successfully complete their daily reports for a week in both countries. The number of social network members was 741 for South Korean adolescents and 760 for German adolescents. The number of social network members corresponded to the size of each sample.

On the basis of her factor analyses on the combined sample of South Korean and German adolescents, Park (personal communication, May 1998) has suggested two potential latent variables underlying the seven social functions. One latent variable, called 'positive function', was associated with five social functions such as attachment, control, guidance, intimacy, and self-validation, and the other, called 'negative function' was associated with two social functions such as comparison and conflict. Hence we assumed a single two latent variable model that was fitted to both samples.

In the model, the latent variables were specified in the same way as above, and they were assumed to have direct influences on trust of adolescents toward other social network members. The specified model is provided in Figure 3.5.

In Figure 3.5, $\mathbf{Z}_j^{(2)}$ was an n_j by 7 matrix ($j = 1,2$), whose first five columns corresponded with the five social functions associated with positive function, whereas the last two corresponded with the two social functions related to negative function. On the other hand, $\mathbf{Z}_j^{(1)}$ was a column vector of order n_j, corresponding to an endogenous variable, trust. In the model, $\mathbf{W}_j$ and $\mathbf{A}_j'$ were specified equally across samples as follows:

$$\mathbf{W}_j = \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ w_3 & 0 \\ w_4 & 0 \\ w_5 & 0 \\ 0 & w_6 \\ 0 & w_7 \end{bmatrix},$$

and

$$\mathbf{A}_j' = \begin{bmatrix} a_1 & a_2 \end{bmatrix}.$$

Using the multi-sample comparison feature in extended redundancy analysis, we simultaneously fit the two latent variable model to South Korean and German adolescent samples in order to examine differences in two sets of parameter estimates.

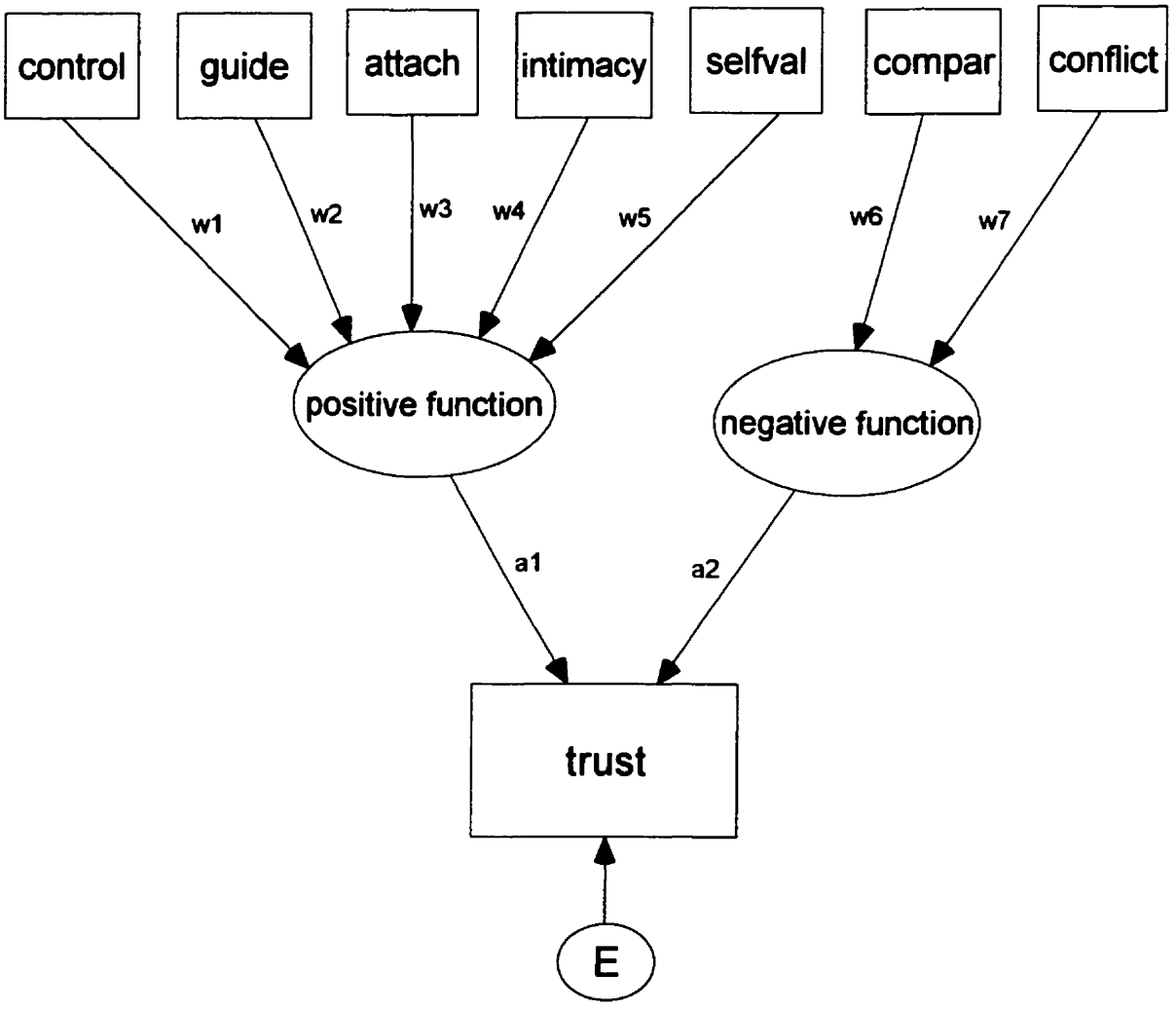

Figure 3.5. The two latent variable model for the social function data.

Results are presented in Figures 3.6a and 3.6b. The bootstrapped standard deviations and critical ratios of parameter estimates in both samples are given in Table 3.3.

The model fit obtained from this multi-sample analysis was .166. It turned out to be significant in terms of its bootstrapped critical ratio (11.9). It was found that in both South Korean and German samples, control, attachment, and intimacy seemed to be significantly associated with positive function. In the South Korean sample, on the other hand, conflict was likely to be significantly related to negative function, whereas in the German sample both comparison and conflict were found to be significantly associated with negative function. In both samples of adolescents, positive function showed a positive and significant influence on trust (i.e., .46 and .42 in South Korea and Germany, respectively), while negative function had a negative and significant influence on trust (i.e., -.09 and -.19 in South Korea and Germany, respectively). It was thus found that for both samples, positive function tended to have an effect of similar strength on trust. On the other hand, negative function in the German sample seemed to show a stronger effect on trust than that in the South Korean sample (i.e., negative function affected trust more negatively). The correlations between two latent variables were .26 and .45 in South Korean and German adolescents, respectively. They turned out to be significant in terms of their bootstrapped critical ratios (2.2 and 9.7 for the South Korean and German samples, respectively). The squared multiple correlations of trust were equal to .19 and .14 in the South Korean and German samples, respectively. Both turned out to be significant according to their bootstrapped critical ratios (8.7 and 8.6 for the South Korean and German samples, respectively).

Fit = .1660

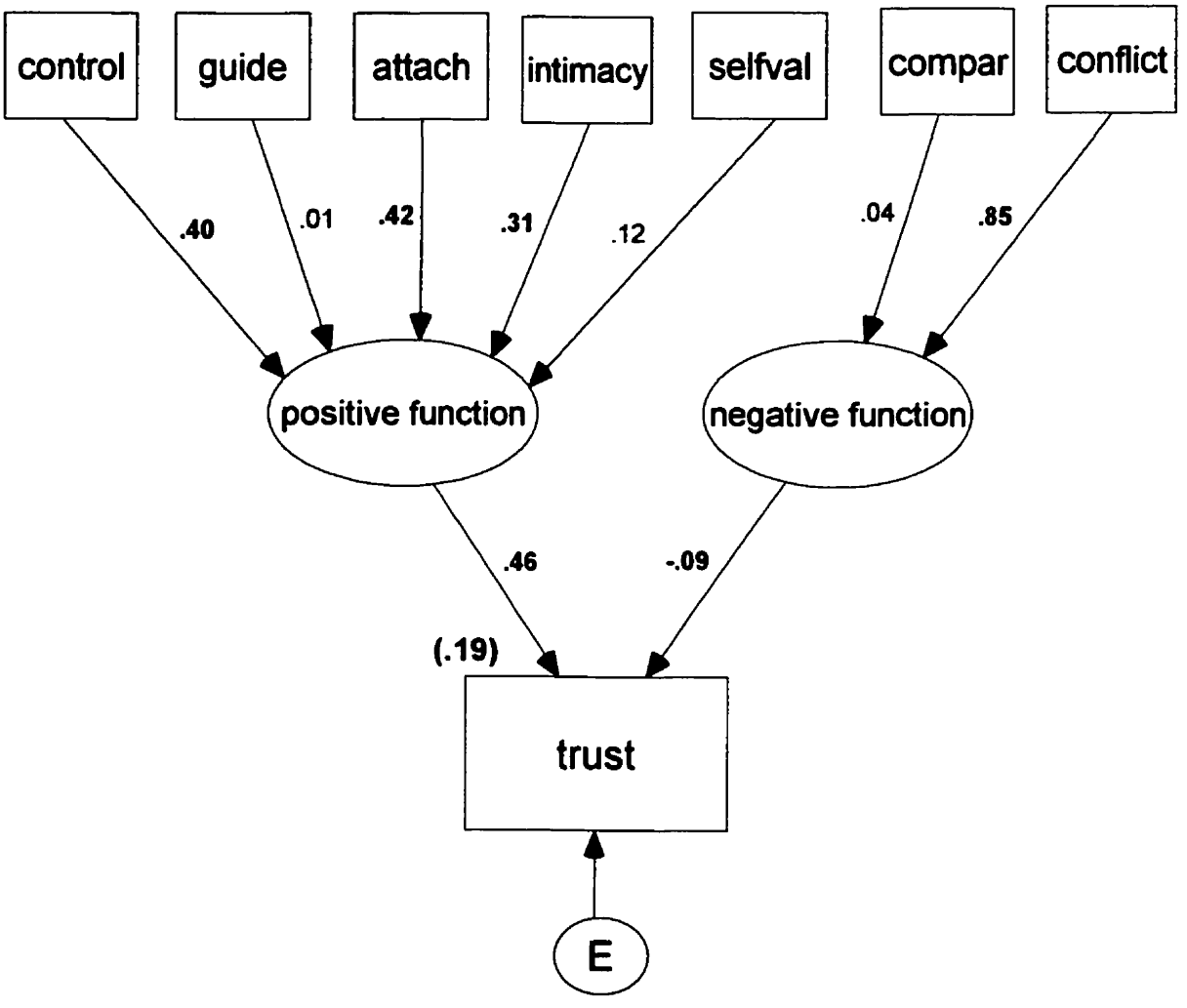

Figure 3.6a. The unconstrained multi-sample analysis for the social function data (Korean adolescents).

Fit = .1660

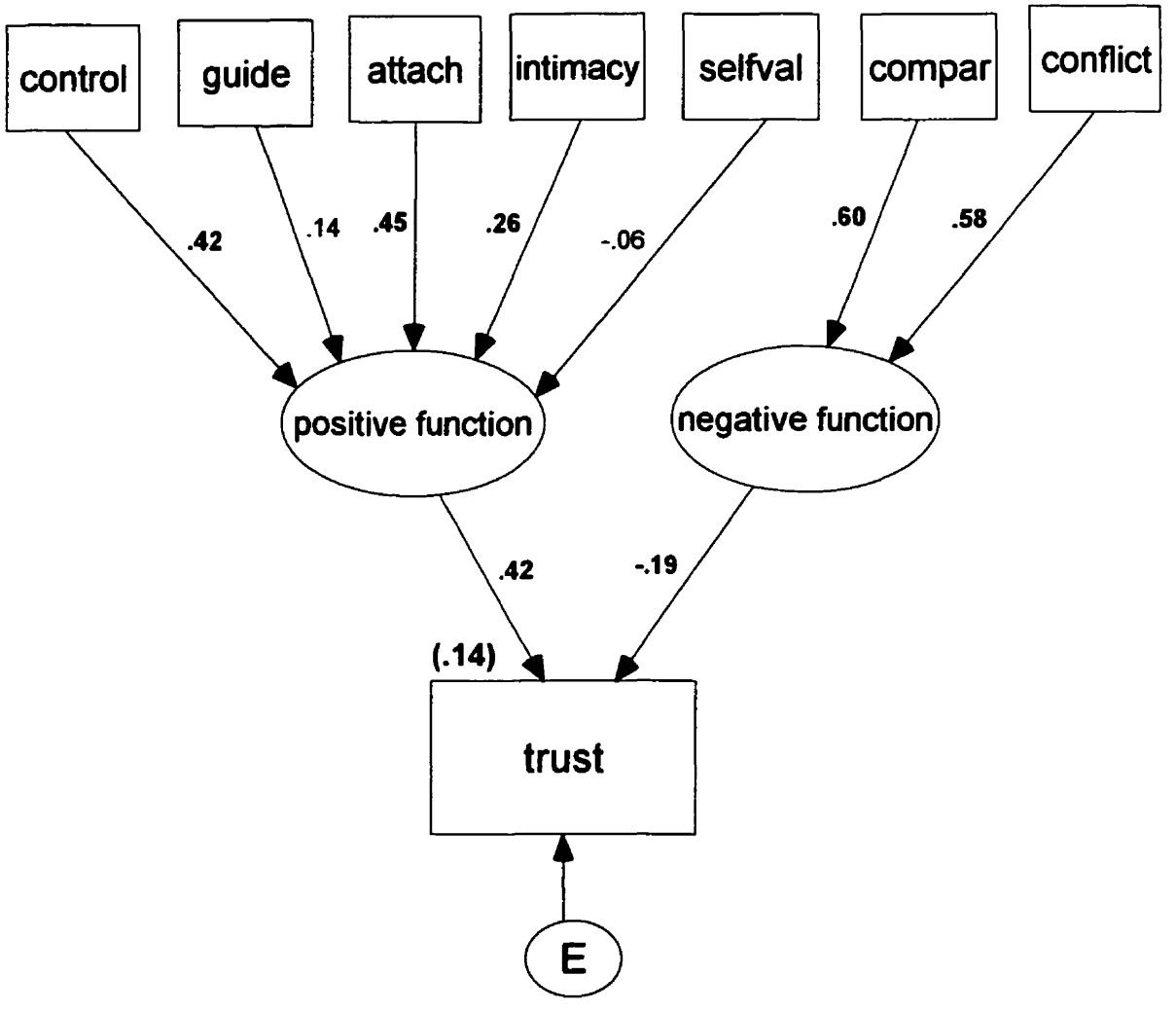

Figure 3.6b. The unconstrained multi-sample analysis for the social function data (German adolescents).

Table 3.3. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the unconstrained multi-sample analysis for the social function data.

		Estimates	SE	CR
KA[1]	w_1	.40	.08	5.0
	w_2	.01	.08	0.1
	w_3	.42	.12	3.5
	w_4	.31	.08	3.9
	w_5	.12	.09	1.3
	w_6	.04	.48	0.1
	w_7	.85	.35	2.4
	a_1	.46	.03	15.3
	a_2	-.09	.03	-3.0
GA[2]	w_1	.42	.10	4.2
	w_2	.14	.09	1.6
	w_3	.45	.11	4.1
	w_4	.26	.13	2.0
	w_5	-.06	.09	-0.7
	w_6	.60	.15	4.0
	w_7	.58	.14	4.1
	a_1	.42	.03	14.0
	a_2	-.19	.04	-4.8

1. KA = Korean Adolescents
2. GA = German Adolescents

On the basis of the obtained solutions in multi-sample analysis, we further assumed that the effects of control and attachment on positive function were identical across two samples. We also hypothesized that positive function had the same effect on trust across two samples. These across-sample assumptions could be incorporated by imposing equality constraints on the portions of free parameters in $\dot{\mathbf{W}}$ and $\mathring{\mathbf{A}}'$, that is, $\dot{\mathbf{w}}^*$ and $\mathring{\mathbf{a}}^*$, as described in the previous section. Let $\mathbf{P}'$ denote a matrix of equality constraints on $\dot{\mathbf{w}}^*$. Let $\mathbf{p}'$ denote a vector of equality constraints on $\mathring{\mathbf{a}}^*$. The $\mathbf{P}'$ and $\mathbf{p}'$ was easily specified as

$$\mathbf{P}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{p}' = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix}.$$

We require $\mathbf{P}'\dot{\mathbf{w}}^* = \mathbf{0}$ and $\mathbf{p}'\mathring{\mathbf{a}}^* = 0$. In the above, the first row of $\mathbf{P}'$ indicated that the component weight for control (corresponding to w_1) was identical across two samples. The second row of $\mathbf{P}'$ represented that the component weight for attachment was equivalent across two samples. Likewise, $\mathbf{p}'$ meant that the loadings for positive function were equal across the two samples. Then the $\mathbf{P}'$ and $\mathbf{p}'$ were transformed into reparametrization forms by the procedure described in Section 2.4. Results of incorporating the equality constraints into the model were provided in Figures 3.7a and 3.7b. The standard deviations and critical ratios of the parameter estimates obtained from the bootstrap method with 100 boostrap samples are presented in Table 3.4.

Fit = .1658

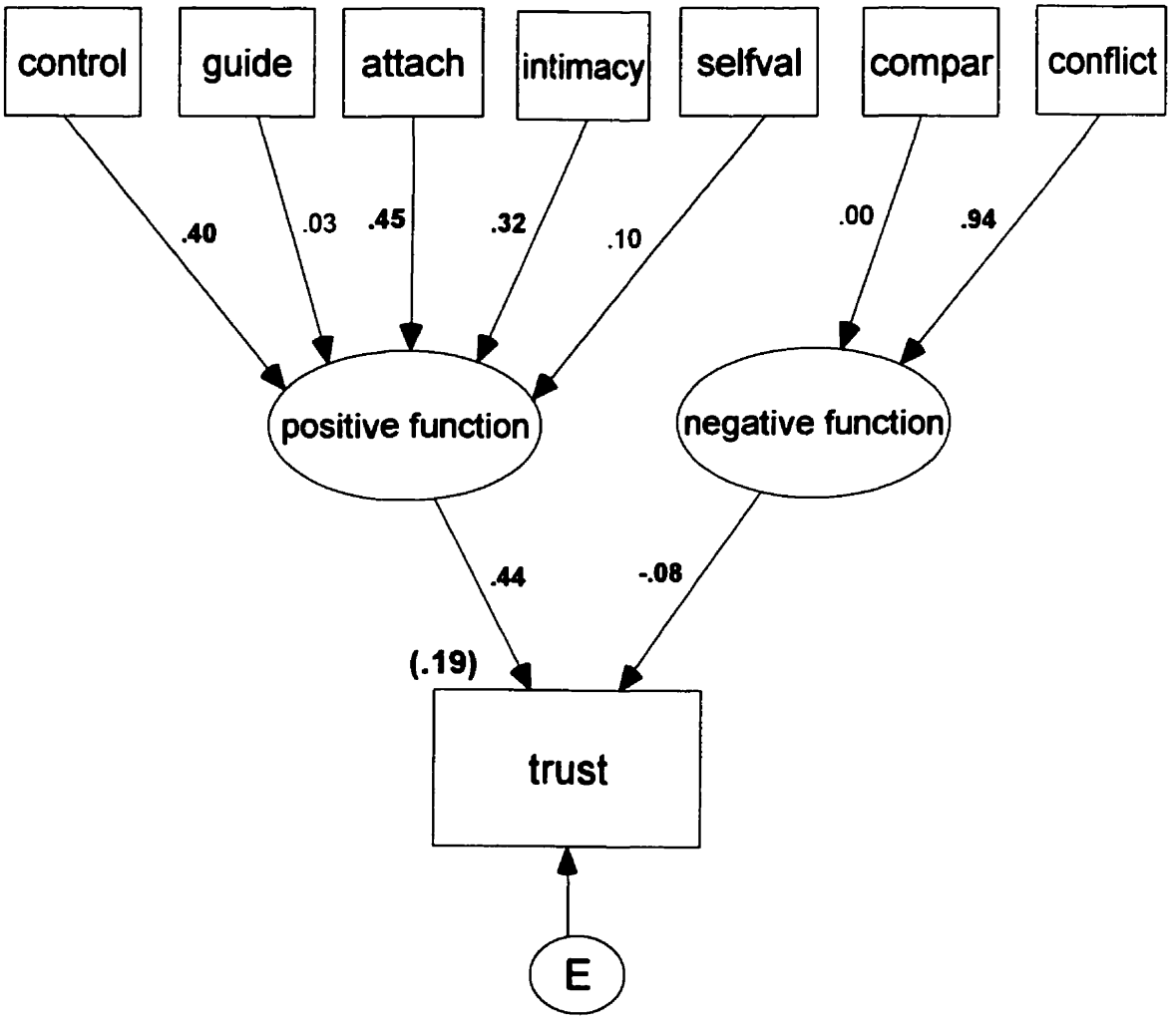

Figure 3.7a. The constrained multi-sample analysis for the social function data (Korean adolescents).

Fit = .1658

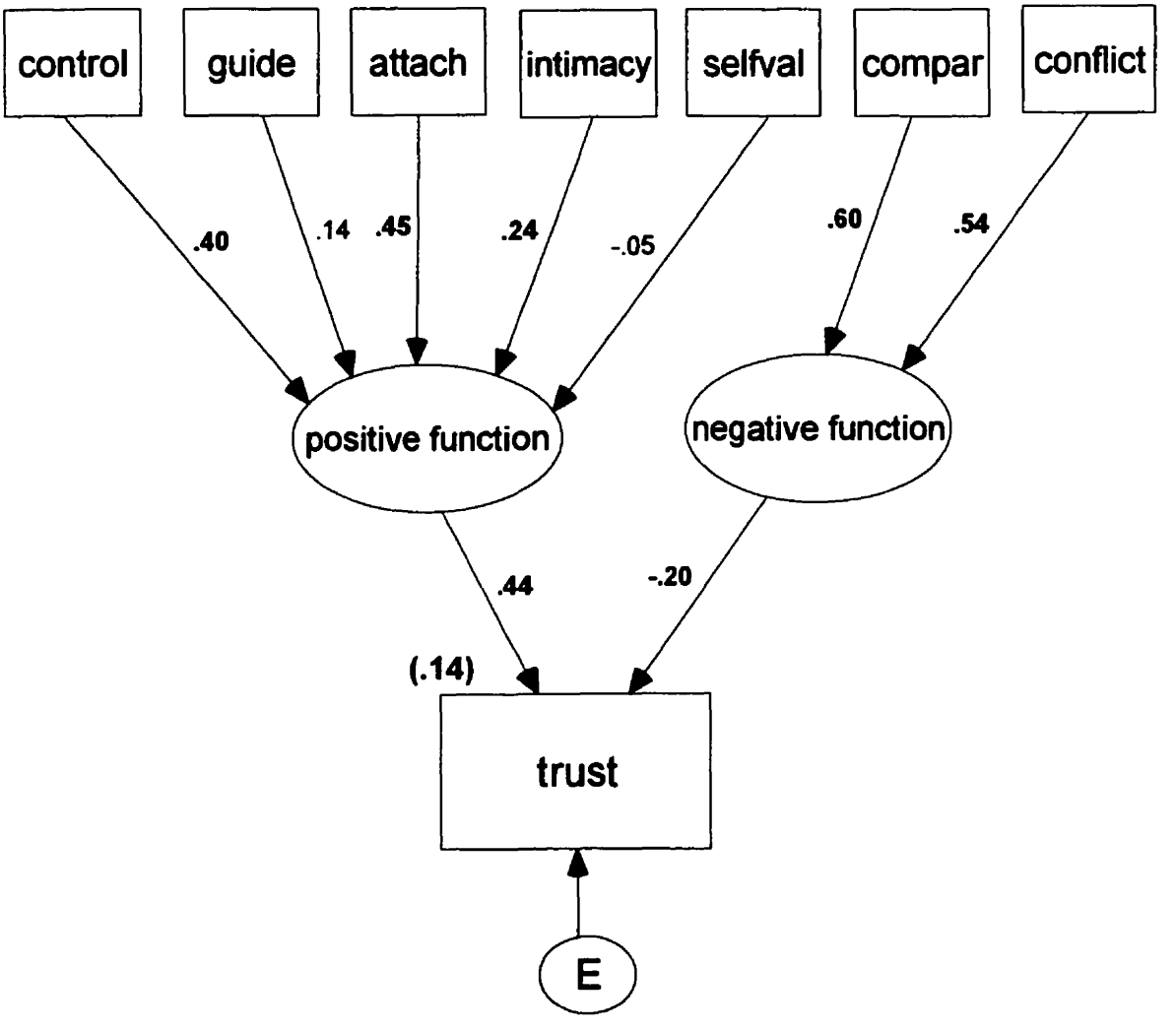

Figure 3.7b. The constrained multi-sample analysis for the social function data (German adolescents).

Table 3.4. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from the constrained multi-sample analysis for the social function data.

		Estimates	SE	CR
KA[1]	w_1	.40	.06	6.7
	w_2	.03	.11	1.6
	w_3	.45	.07	6.4
	w_4	.32	.08	4.0
	w_5	.10	.10	1.0
	w_6	.00	.47	0.0
	w_7	.94	.28	3.4
	a_1	.44	.02	22.0
	a_2	-.08	.03	-2.7
GA[2]	w_1	.40	.06	6.7
	w_2	.14	.08	1.8
	w_3	.45	.07	6.4
	w_4	.24	.09	2.7
	w_5	-.05	.09	-0.6
	w_6	.60	.15	4.0
	w_7	.54	.16	3.4
	a_1	.44	.02	22.0
	a_2	-.20	.04	-5.0

1. KA = Korean Adolescents
2. GA = German Adolescents

In the figures, due to the equality constraints imposed, the component weight of control was identical across two samples (.40) and also turned out to be significant. This was the case for the component weight of attachment (.45) and the loading for positive function (.44). The goodness of fit of the sample-wise constrained model was .1658. It was almost the same fit as that obtained from the unconstrained multi-sample analysis. The fit turned out to be significant (the bootstrapped critical ratio = 10.8). Also, the constrained multi-sample analyses provided quite similar solutions to the unconstrained case. Therefore, it might be safe to say that our hypotheses regarding the relationships among the parameters across samples were reasonable. This yields simpler interpretations of the solutions, reducing the number of parameters to be interpreted.

Chapter 4

Further Considerations

In this chapter, we discuss a number of topics that further enhance the capabilities of the proposed method. They include data transformations, missing data, metric matrices, robust estimation, and efficient estimation. Illustrative examples are provided to demonstrate how they are handled in the proposed method.

4. 1. Data Transformations

In structural equation models, all observed variables are usually assumed to be numerical or measured at an interval or ratio scale level. However, our method can readily analyze categorical variables through data transformations. In this section, we are concerned with a certain type of data transformations, often called optimal scaling. In optimal scaling, variables are assumed to be related to their model predictions, and then are transformed in such a way that they agree with the model predictions as much as possible, while the measurement characteristics assumed of variables are strictly maintained. A variety of measurement restrictions can be imposed on the transformation, depending on the measurement characteristics of the data, such as measurement levels (nominal, ordinal, or numerical) and measurement process (discrete or continuous). For instance, for an ordinal variable, it is required that its

observations should be order preserving and tied observations should remain tied (in the case of discrete ordinal variables) or become untied (in the case of continuous ordinal variables). The transformation under this restriction is called a monotonic transformation (e.g., Kruskal, 1964; Ramsay, 1988, 1998). Refer to Young (1981) for detailed information on the measurement restrictions.

On the other hand, model parameters are estimated so that the model predictions derived from the parameter estimates are as close as the (transformed) variables as possible under the constraints (e.g., normalization restrictions, orthogonality restrictions, etc.) imposed on those parameters. See van Buuren (1990, Chapter 3) for more varieties of parameter restrictions.

A number of such data transformations have been proposed in redundancy analysis (e.g., Israëls, 1984, Meulman, 1986, van der Burg, 1988, etc.). We may follow a similar approach: briefly speaking, we view the data matrices, $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$, as data parameter matrices, denoted by $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, respectively. The data parameters are 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; e.g., some variables can be nominal, others ordinal, and others interval. This indicates 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.

More specifically, the ALS procedure with the data transformation feature

proceed 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, our problem amounts to minimizing

$$\begin{aligned} f &= \mathrm{SS}(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\mathbf{W}\mathbf{A}') \\ &= \mathrm{SS}(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\mathbf{B}), \end{aligned} \tag{4.1}$$

with the conditions that $\mathrm{diag}(\mathbf{W}'\mathbf{S}^{(2)\prime}\mathbf{S}^{(2)}\mathbf{W}) = \mathbf{I}$, $\mathbf{s}_i'\mathbf{s}_i = 1$, and $\mathbf{s}_i = \xi(\mathbf{z}_i)$, where ξ refers to a transformation of the observations in $\mathbf{z}_i$, which is a function of their measurement characteristics.

To minimize (4.1), 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 represents estimating $\mathbf{W}$ and $\mathbf{A}'$, for fixed $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, which is analogous to the procedure described in Chapter 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 (4.1). 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 $\hat{\mathbf{s}}_i$ denote the model prediction of $\mathbf{s}_i$. Then (4.1) may be rewritten as

$$f = \sum_{i=1}^{p+q} \mathrm{SS}(\mathbf{s}_i\boldsymbol{\eta}' - (\boldsymbol{\Delta} - \boldsymbol{\Psi})). \tag{4.2}$$

In (4.2), $\boldsymbol{\eta}'$, $\boldsymbol{\Delta}$, 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,

$$\boldsymbol{\Delta} = \begin{cases} \mathbf{S}^{(2)}_{(h)}\mathbf{B}_{(h)} & \text{if } \mathbf{s}_i \text{ is shared} \\ \mathbf{S}^{(2)}\mathbf{B} & \text{otherwise} \end{cases},$$

$$\boldsymbol{\Psi} = \mathbf{S}^{(1)}_{(g)},$$

$$\boldsymbol{\eta}' = \begin{cases} \mathbf{e}_g' - \mathbf{b}_h' & \text{if } \mathbf{s}_i \text{ is shared} \\ \mathbf{e}_g' & \text{otherwise} \end{cases}.$$

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

$$\boldsymbol{\Delta} = \mathbf{S}^{(2)}_{(h)}\mathbf{B}_{(h)},$$

$$\boldsymbol{\Psi} = \mathbf{S}^{(1)},$$

$$\boldsymbol{\eta}' = \mathbf{b}_h'.$$

In the above, matrix $\mathbf{S}^{(2)}_{(h)}\mathbf{B}_{(h)}$ 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 $\mathbf{S}^{(1)}_{(g)}$ 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, $\hat{\mathbf{s}}_i$ is obtained by

$$\hat{\mathbf{s}}_i = \mathbf{\Lambda\eta}(\mathbf{\eta}'\mathbf{\eta})^{-1}, \tag{4.3}$$

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

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

$$\mathbf{s}_i = \mathbf{\Upsilon}_i(\mathbf{\Upsilon}_i'\mathbf{\Upsilon}_i)^{-1}\mathbf{\Upsilon}_i'\hat{\mathbf{s}}_i. \tag{4.4}$$

In (4.4), $\mathbf{\Upsilon}_i$ is determined by the measurement restrictions imposed on the transformation. For example, for nominal variables, $\mathbf{\Upsilon}_i$ is an indicator matrix, whose element stands for category membership, and is known in advance. For ordinal variables, on the other hand, $\mathbf{\Upsilon}_i$ 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 this step, $\mathbf{s}_i$ may also be transformed by a maximum likelihood method (e.g., Ramsay, 1988, 1998; Takane, 1978; Winsberg & Ramsay, 1980, 1983). Box and Cox

(1964) recognized that the least squares transformation of dependent variables was likely to lead to serious bias problems, and instead the maximum likelihood or the Bayesian method which took into account the Jacobian of the transformation (i.e., the determinant of the first derivative of the transformation with respect to variables) might be preferred (also see Ramsay, 1988). In the case of monotonic transformations, smoothness of the transformation can be of virtue (Ramsay, 1998), since it plays a role in stabilizing estimated transformations. By contrast, the least squares monotonic transformation often looks like a step function. In such cases, smooth monotonic transformations based on the maximum likelihood method (e.g., Ramsay, 1988, 1998) can be employed. In particular, Ramsay (1998) developed a computationally elegant procedure for estimating smooth monotonic transformations, which amounts to estimating an arbitrary twice differentiable strictly monotone function defined on an interval $[0, \infty)$, maximizing a penalized (log) likelihood criterion by the Newton-Raphson method. For the transformation of each variable, the fitting criterion may be written as

$$f = n^{-1}\mathrm{SS}(\hat{\mathbf{s}}_i - \alpha_i - \beta_i\, m(\mathbf{z}_i)) \; + \lambda_i \int \omega^2(\mathbf{z}_i)\mathrm{d}\mathbf{z}_i, \tag{4.5}$$

where α_i and β_i are regression coefficients, $m(\mathbf{z}_i) = \{D^{-1}\exp(D^{-1}\omega)\}(\mathbf{z}_i)$ (D^{-1} refers to the partial integration operator), and λ_i is the smoothing parameter determining the amount of penalty. Function ω may be defined as a linear combination of some set of basis functions ϕ_v, $v = 1, \cdots, V$. Let $\boldsymbol{\phi} = [\phi_1, \cdots, \phi_V]'$. Then, $\omega(\mathbf{z}_i) = \boldsymbol{\tau}_i'\boldsymbol{\phi}(\mathbf{z}_i)$, where $\boldsymbol{\tau}_i$ is the coefficient vector defining the linear combination. Criterion (4.5) is thus

minimized with respect to τ_i, α_i, and β_i in order to update $\mathbf{s}_i$.

In (4.2), we see that updating a variable is dependent on other variables. To assure convergence, therefore, 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 $\mathbf{S}^{(2)}$. This indicates that at each iteration, the number of variables being updated is equal to $p + q^*$, where $q^* = q -$ the number of common variables in $\mathbf{S}^{(2)}$.

The proposed method is akin to those used in redundancy analysis in that data transformations are incorporated by the process of alternating two main phases until convergence is obtained. In our procedure, the model parameters are iteratively re-estimated due to the structure of $\mathbf{W}$ and $\mathbf{A}'$, whereas they are analytically solved in redundancy analysis. In addition, the data transformation phase in our method turns out to be a bit more complicated than its counterparts in redundancy analysis, in order to take into account the variables common to both $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, which frequently appear in structural equation models.

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 (Takane & Shibayama, 1991). This also enables 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 attainable since it asks for a well-defined global criterion that is consistently minimized/maximized by updating the transformed variables.

Categorical variables can also be analyzed in covariance structure analysis (e.g., Muthén, 1978, 1984, 1987). Covariance structure analysis cannot be directly applied to categorical variables, since they are usually non-normal, and their population covariance structure cannot be specified by model parameters. It is also difficult to use the ADF estimator since categorical variables can also yield heteroscedastic errors and the ADF estimator assumes homoscedasticity (Bollen, 1989). Instead, it is assumed that there exists a latent continuous variable, a so-called response strength, underlying an observed categorical variable, and both variables are connected by a threshold model (i.e., the observed variable is discretized from the latent continuous variable). Then under the assumption of normality of the latent continuous variable, correlations among the latent continuous variables (e.g., the polychoric correlation for two ordinal observed variables and the tetrachoric correlation for two dichotomous observed variables) are calculated based on the discretization thresholds, and the differences between the sample correlations and implied correlations are minimized by the generalized least squares or unweighted least squares method. However, the assumption that the underlying latent continuous variable is normal may be often invalid. In addition, correlations among the latent variables are only measured, and improper solutions can occur and factor scores are indeterminate. Examples of the least squares monotonic transformation and the smooth monotonic transformation will be provided in Section

4.6.

4. 2. Missing Data

It is not uncommon that some observations are missing for some reasons, particularly in large data sets. A number of methods have been proposed to deal with missing data. One simple procedure for handling missing values is to delete any cases having at least one missing observation. However, this is unsatisfactory if missing values are numerous and scattered throughout the data set, as deletion of the cases may incur substantial loss of information. Another method is to estimate missing values prior to analysis and then use the estimates in subsequent data analysis. We may, for example, use the mean of non-missing values or any prior knowledge/experience to replace a missing value by some actual value.

A third method is to iteratively re-estimate values for missing observations (e.g., Gabriel & Zamir, 1979; Gifi, 1990). We start by completing the data with some initial estimates for missing data, obtain model estimates by fitting the model to the complete data, update estimates of missing values based on the model estimates, fit the model to the updated data, and so on. These procedures are repeated until no significant changes take place in the estimates. This approach may be particularly advantageous in the context of our method, since it can be readily incorporated into the ALS procedure described in the previous section (e.g., Young, de Leeuw, Takane, 1976). That is, if some values in $\mathbf{s}_i$ are missing, we start by filling in the mean of non-missing values. Then, we estimate $\mathbf{s}_i$, and only the non-missing values are transformed based on their corresponding values in $\hat{\mathbf{s}}_i$ while the missing values are simply replaced by the

corresponding values in $\hat{\mathbf{s}}_i$. The updated $\mathbf{s}_i$ consisting of the newly transformed non-missing and re-estimated missing values is then normalized, and is used as a substitute for the previous $\mathbf{s}_i$.

This process, however, may be computationally expensive since missing values are variable-wise estimated, that is, missing values in a single variable are only handled at a time. Furthermore if a certain type of data transformation is applied for non-missing values, additional substeps are required to distinguish missing parts and non-missing parts in $\mathbf{s}_i$, which are treated differently.

If no data transformations are involved, we may consider a different ALS approach to iteratively estimate missing values. In this approach we view missing values in $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ as additional parameters, and they are estimated with non-missing values fixed. To estimate the missing values in $\mathbf{Z}^{(2)}$ for fixed $\mathbf{W}$, $\mathbf{A}'$, and $\mathbf{Z}^{(1)}$, (2.8) may be written as

$$\begin{aligned} f &= \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - \mathrm{vec}(\mathbf{Z}^{(2)}\mathbf{B})) \\ &= \mathrm{SS}(\mathrm{vec}(\mathbf{Z}^{(1)}) - (\mathbf{B}'\otimes \mathbf{I})\mathrm{vec}(\mathbf{Z}^{(2)})). \end{aligned} \tag{4.6}$$

Let $\Xi = \mathbf{B}'\otimes \mathbf{I}$. Also, let $\mathbf{z}^{*(2)}$ denote the vector formed by eliminating non-missing values from $\mathrm{vec}(\mathbf{Z}^{(2)})$. Let Ξ^* denote the matrix formed by eliminating the columns of Ξ corresponding to the non-missing values in $\mathrm{vec}(\mathbf{Z}^{(2)})$. Then, the least squares estimate of missing values in $\mathbf{Z}^{(2)}$, say $\hat{\mathbf{z}}^{*(2)}$, is obtained by

$$\hat{\mathbf{z}}^{*(2)} = (\Xi^{*\prime}\Xi^{*})^{-1}\Xi^{*\prime}\mathrm{vec}(\mathbf{Z}^{(1)}), \tag{4.7}$$

assuming that $\Xi^{*\prime}\Xi^{*}$ is nonsingular. $\mathrm{vec}(\mathbf{Z}^{(2)})$ is easily updated from $\hat{\mathbf{z}}^{*(2)}$, and is

normalized. vec($\mathbf{Z}^{(1)}$) is then updated by simply replacing the previous estimates of missing values with the corresponding elements of vec($\mathbf{Z}^{(2)}\mathbf{B}$)), and is also normalized. These procedures are repeated until convergence occurs. When some variables are shared by $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$, they should remain identical during iterations. This may be done by setting the shared variable in $\mathbf{Z}^{(1)}$ equal to the updates of the same variable in $\mathbf{Z}^{(2)}$. This approach seems to be computationally efficient because all missing values in a data matrix are estimated in a single step.

4. 3. Metric Matrices

Our estimation procedure can incorporate metric matrices in the optimization criterion. Two kinds of metric matrices may be considered, one on the row side, and the other on the column side of $\mathbf{Z}^{(1)}$ (e.g., Takane & Shibayama, 1991). Let $\mathbf{K}$ denote an n by n row-side metric matrix. Let $\mathbf{L}$ denote a p by p column-side metric matrix. Matrices $\mathbf{K}$ and $\mathbf{L}$ are both assumed to be non-negative definite. To estimate $\mathbf{W}$ and $\mathbf{A}'$ with these metric matrices incorporated, we minimize

$$f = \mathrm{SS}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')_{K,L}, \tag{4.8}$$

where $\mathrm{SS}(\mathbf{X})_{KL} = \mathrm{trace}(\mathbf{KXLX}')$. When $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$, (4.8) reduces to (2.8). Even when $\mathbf{K} \neq \mathbf{I}$ and/or $\mathbf{L} \neq \mathbf{I}$, however, (4.8) can be reduced to (2.8) by a simple transformation (e.g., Rao, 1980). Let $\mathbf{K} = \mathbf{R}_K\mathbf{R}_K'$ and $\mathbf{L} = \mathbf{R}_L\mathbf{R}_L'$ be any square root decompositions of $\mathbf{K}$ and $\mathbf{L}$. Then, (4.8) may be rewritten as

$$f = \mathrm{SS}(\mathbf{R}_K'\mathbf{Z}^{(1)}\mathbf{R}_L - \mathbf{R}_K'\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}'\mathbf{R}_L)$$
$$= \mathrm{SS}(\tilde{\mathbf{Z}}^{(1)} - \tilde{\mathbf{Z}}^{(2)}\mathbf{W}\tilde{\mathbf{A}}'), \qquad (4.9)$$

where $\tilde{\mathbf{Z}}^{(1)} = \mathbf{R}_K'\mathbf{Z}^{(1)}\mathbf{R}_L$, $\tilde{\mathbf{Z}}^{(2)} = \mathbf{R}_K'\mathbf{Z}^{(2)}$, and $\tilde{\mathbf{A}}' = \mathbf{A}'\mathbf{R}_L$. This is essentially the same as (2.8), and may be minimized in a similar way. In (4.9), we can estimate $\mathbf{A}'$ without destroying its structure (e.g., zero elements), because the least square estimate of $\mathbf{a}$ is obtained by

$$\hat{\mathbf{a}} = (\mathbf{\Phi}^{*\prime}\mathbf{\Phi}^{*})^{-1}\mathbf{\Phi}^{*\prime}\mathrm{vec}(\tilde{\mathbf{Z}}^{(1)}), \qquad (4.10)$$

where $\mathbf{\Phi}^{*}$ is the matrix formed by eliminating the columns of $\mathbf{\Phi}$ ($= \mathbf{R}_L' \otimes \tilde{\mathbf{Z}}^{(2)}\mathbf{W}$) corresponding to zero elements in $\mathrm{vec}(\mathbf{A}')$.

Judicious choice of metric matrices broadens the capacity of our method. Takane and Hunter (2000) provided examples of non-identity metric matrices used for various purposes. For instance, in Section 2.2, the data are assumed to be *a priori* standardized so as to avoid incomparable scales across different variables. This is equivalent to using the inverse of the diagonal matrix of sample variances of $\mathbf{Z}^{(1)}$ as $\mathbf{L}$. Meredith and Millsap (1985) proposed to use the matrix of reliability coefficients or of inverses of the variances of anti-images (Guttman, 1953) as $\mathbf{L}$. When the columns of the residual matrix are correlated and/or have markedly different variances after a model is fitted to the data, the variance-covariance matrix among the residuals may be estimated, and its inverse be used as $\mathbf{L}$. This has the effect of getting smaller expected mean squared errors of parameter estimates by orthonormalizing the residuals in evaluating the overall goodness of fit (Takane & Hunter, 2000). In maximum likelihood common factor

analysis, scale invariance is obtained by scaling the data (with communalities on the diagonal) by Θ^{-1}, where Θ^2 is the diagonal matrix of uniqueness. This is the same as setting $\mathbf{L} = \Theta^{-2}$ in our method, assuming that Θ^2 is known in advance. There are a number of methods proposed to estimate Θ^2 non-iternatively (e.g., Ihara & Kano, 1986). In addition, Rao (1964, Section 9) argued that scale invariance could be achieved by specifying certain non-identity $\mathbf{L}$ matrices.

When rows of $\mathbf{Z}^{(1)}$ consist of several responses made by the same subject, they are likely to be correlated. In this case, a matrix of serial correlations is estimated, and its inverse can be used as $\mathbf{K}$ (Escoufier, 1987) in order to achieve independence among observations. When differences in importance and/or in reliability among the rows are suspected, a special kind of diagonal matrix may be used for $\mathbf{K}$ that has the effect of differentially weighting rows of a data matrix. (In correspondence analysis, for example, the square root of row totals of a contingency table is used as $\mathbf{K}$.)

Although $\mathbf{K}$ and $\mathbf{L}$ are specified as fixed matrices above, we may also use iteratively updated metric matrices during the optimization procedure. Such metric matrices may be considered for robust estimation, which is presented in the next section.

4. 4. Outlier Diagnostics and Robust Estimation

Our method may not be robust against outliers as far as it is based on solving ordinary least squares (OLS) criterion, which amounts to minimizing the sum of the 'squared' residuals of (2.8). In general, there exist two approaches to deal with outliers: Outlier diagnostics and Robust estimation. In the diagnostic approach, we try to identify

outliers, remove or adjust them, and fit the data by a traditional method. For instance, leverage points may be used to assess influence of each row in $\mathbf{Z}^{(2)}$. The leverage point of the ith row ($i = 1, \cdots, n$), denoted by h_{ii}, is equivalent to the ith diagonal element of $\mathbf{Z}^{(2)}(\mathbf{Z}^{(2)\prime}\mathbf{Z}^{(2)})^{-1}\mathbf{Z}^{(2)\prime}$. Influential points are usually determined by $h_{ii} > 2q/n$ or $3q/n$. Mahalanobis distance (MD) is another useful diagnostic tool to detect influence of rows in $\mathbf{Z}^{(2)}$. The MD for the ith row is calculated by

$$\mathrm{MD}_i^2 = (\mathbf{z}_i^{(2)} - \bar{z}_i^{(2)}\mathbf{1}_q)\mathbf{C}^{-1}(\mathbf{z}_i^{(2)} - \bar{z}_i^{(2)}\mathbf{1}_q)',$$

where $\mathbf{z}_i^{(2)}$ is the ith row of $\mathbf{Z}^{(2)}$, $\bar{z}_i^{(2)}$ is the average of the ith row, $\mathbf{C}$ is the covariance matrix of $\mathbf{Z}^{(2)}$, and $\mathbf{1}_q$ is the q-component vector of ones. The MD has the same diagnostic power as h_{ii} since

$$\mathrm{MD}_i^2 = (n-1)(h_{ii} - 1/n).$$

To examine the influence of the ith row in $\mathbf{Z}^{(1)}$, we may use Cook's squared distance (CD^2), defined as

$$\mathrm{CD}^2(i) = \{[\hat{\mathbf{z}}_i^{(1)} - \hat{\mathbf{z}}_i^{(1)}(i)]\mathbf{M}[\hat{\mathbf{z}}_i^{(1)} - \hat{\mathbf{z}}_i^{(1)}(i)]'\}/\varepsilon,$$

where $\hat{\mathbf{z}}_i^{(1)}$ is the ith row of the LS estimates from full $\mathbf{Z}^{(1)}$, $\hat{\mathbf{z}}_i^{(1)}(i)$ is the ith row of the LS estimates from $\mathbf{Z}^{(1)}$ with the ith row deleted, and $\mathbf{M} = \mathbf{Z}^{(2)\prime}\mathbf{Z}^{(2)}$. Constant ε is usually chosen as $\varphi[\sum r_i^2/(n-\varphi)]$, where $\sum r_i^2$ is the sum of residuals, and $\varphi = \sum h_{ii}$. The large value of $\mathrm{CD}^2(i)$ implies that the ith row is influential. For more detailed expositions of diagnostics in regression, refer to Belsley, Kuh, & Welsch (1980), Cook & Weisberg (1982), etc.

In the robust estimation approach, on the other hand, we start by fitting a model to a majority of data points, and detect outliers which are observations with large residuals. Rousseeuw and Leroy (1987) pointed out that the two approaches pursued the same goal but proceeded in opposite directions. Nonetheless the classical diagnostics based on OLS often fail to detect multiple outliers (e.g., Singh, 1996; Walczak & Massart, 1995).

Griep, Walkeling, Vankeerberghen, and Massart (1995) applied three different robust estimation methods (i.e., least median of squares, Siegel's repeated median, and Beaton and Tukey's iteratively reweighted least squares) to PLS. To evaluate their performance, they were applied to a data set with outliers artificially added, and the solutions were compared with the ordinary PLS solutions without outliers. It was found that the iteratively reweighted least squares (IRLS) method (Beaton & Tukey, 1974) performed better at least for the low dimensional PLS than the other methods. The least median of squares (LMS) is known as the best in regression analysis (Rousseeuw & Leroy, 1987), which requires the slope as well as the intercept to calculate a set of squared residuals. In PLS, however, data sets are preprocessed to be mean-centered, so that there are no intercepts available. Griep *et al.* (1995) pointed out that this might influence the performance of LMS. Since in our method data matrices are usually standardized in advance and a few linear components are extracted from the data matrices, we may employ IRLS to handle outliers.

In IRLS, a data weight matrix (different from the component weight matrix, **W**) is obtained to give different weights to observations, depending on the size of outliers.

The data weight matrix is iteratively re-estimated by the following procedure. Let **U** denote the l by l diagonal matrix of data weights, where $l = n \times p$. That is, **U** is

$$\text{diag}(u_{11}, u_{22}, \cdots, u_{ll}).$$

Let $\mathbf{U} = \mathbf{R}_U \mathbf{R}_U'$. Then, the fitting criterion may be written as

$$\begin{aligned} f &= \text{SS}((\text{vec}(\mathbf{Z}^{(1)}) - \text{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}'))_{U,I} \\ &= \text{SS}(\mathbf{R}_U'(\text{vec}(\mathbf{Z}^{(1)}) - \text{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')). \end{aligned} \tag{4.11}$$

In (4.11), **W** and $\mathbf{A}'$ can be obtained for fixed **U** by the same ALS procedure as in the previous section. (Matrix **U** may be initialized by $\mathbf{I}_l$.) After **W** and $\mathbf{A}'$ are estimated, a vector of residuals, say **r**, is calculated by

$$\mathbf{r} = \text{vec}(\mathbf{Z}^{(1)}) - \text{vec}(\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}'). \tag{4.12}$$

The median of the absolute values of residuals, say δ, is then computed as

$$\delta = \text{median}(|r_i|), \text{ for } i = 1, \cdots, l. \tag{4.13}$$

From (4.12) and (4.13), the elements of **U** are calculated by

$$u_{ii} = \begin{cases} [1 - (r_i/c\delta)^2]^2 & \text{for } |r_i| < c\delta \\ 0 & \text{for } |r_i| < c\delta \end{cases}. \tag{4.14}$$

With the updated **U** fixed, **W** and $\mathbf{A}'$ are re-estimated, then **U** is updated, and so on. These procedures are repeated until convergence is reached. In (4.14), c represents a variable sensitivity factor that defines a threshold beyond which a weight of zero is assigned to that observation. Either 6 or 9 is usually used for c, but the results are quite

similar (Wakeling & Macfie, 1992).

We notice that the above procedure for robust estimation can be viewed as specifying a non-identity and iteratively updated (row-side) metric matrix, **U**. The metric matrix has the effect of differentially weighting rows or observations of data matrices.

4. 5. Efficient Estimation

The assumption of normality is not essential in our method due to the least squares fitting of the ERA model. If it is assumed, nonetheless, we have

$$\begin{aligned} \mathbf{Z}^{(1)} &= \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E} \\ &= \mathbf{Z}^{(2)}\mathbf{B} + \mathbf{E}, \end{aligned} \tag{4.15}$$

and

$$\text{vec}(\mathbf{E}) \sim N(\mathbf{0}, \mathbf{I} \otimes \boldsymbol{\Sigma}), \tag{4.16}$$

where $\boldsymbol{\Sigma}$ is the unknown population covariance matrix of p variables. We may estimate parameters of the ERA model under this specific distributional assumption. In this case, we seek to maximize the log of the likelihood function:

$$g(\mathbf{B}, \boldsymbol{\Sigma}) = \rho + \frac{n}{2}\log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\text{tr}[(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{B})'], \tag{4.17}$$

where $\rho = -\frac{1}{2}np\log(2\pi)$ (e.g., Anderson, 1951, Izenman, 1975, Tso, 1981, Davies & Tso, 1982; Reinsel & Velu, 1998; van der Leeden, 1990).

To maximize (4.17), we may use an optimization procedure similar to an alternating maximum likelihood (AML) procedure (de Leeuw, 1989; van der Leeden,

1990). The procedure consists of two main steps: in the first step, (4.17) is optimized over **B** (that is, **W** and $\mathbf{A}'$), for fixed $\boldsymbol{\Sigma}$. In the second step, (4.17) is optimized over $\boldsymbol{\Sigma}$, for fixed **B**. These steps are alternated until convergence is obtained.

The first step amounts to minimizing

$$g(\mathbf{B} \mid \boldsymbol{\Sigma}) = \frac{1}{2}\mathrm{tr}[(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})'], \tag{4.18}$$

for fixed $\boldsymbol{\Sigma}$. Minimizing (4.18) reduces to minimizing

$$\begin{aligned} g^* &= \mathrm{SS}(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})_{I,\boldsymbol{\Sigma}^{-1}} \\ &= \mathrm{SS}((\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})\mathbf{R}_{\boldsymbol{\Sigma}}) \end{aligned} \tag{4.19}$$

where $\boldsymbol{\Sigma}^{-1} = \mathbf{R}_{\boldsymbol{\Sigma}}\mathbf{R}_{\boldsymbol{\Sigma}}'$. Then, (4.19) can be minimized by the same procedure as given in Section 4.3.

In the next step, we update $\boldsymbol{\Sigma}$ for fixed **B**. This amounts to maximizing

$$g(\boldsymbol{\Sigma} \mid \mathbf{B}) = \rho + \frac{n}{2}\log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\mathrm{tr}[\mathbf{V}\boldsymbol{\Sigma}^{-1}], \tag{4.20}$$

where $\mathbf{V} = (\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})'(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{B})$. Criterion (4.20) may be re-expressed as

$$g^{**}(\boldsymbol{\Sigma} \mid \mathbf{B}) = \rho - \frac{1}{2}(n\log|\boldsymbol{\Sigma}| + \mathrm{tr}[\mathbf{V}\boldsymbol{\Sigma}^{-1}]), \tag{4.21}$$

due to $n\log|\boldsymbol{\Sigma}^{-1}| = n\log|\boldsymbol{\Sigma}|^{-1} = -n\log|\boldsymbol{\Sigma}|$. Maximizing (4.21) amounts to minimizing

$$\tilde{g}^{**} = n\log|\boldsymbol{\Sigma}| + \mathrm{tr}[\mathbf{V}\boldsymbol{\Sigma}^{-1}]. \tag{4.22}$$

By the standard results of matrix derivatives that $\mathrm{dtr}(\mathbf{V}\boldsymbol{\Sigma}^{-1}) = \mathrm{tr}(\mathrm{d}\mathbf{V}\boldsymbol{\Sigma}^{-1})$, $\mathrm{d}(\log|\boldsymbol{\Sigma}|) =$

$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathrm{d}\boldsymbol{\Sigma})$, and $\mathrm{d}(\boldsymbol{\Sigma}^{-1}) = -\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}$ (e.g., Schott, 1997, pp. 332-333),

$$\frac{\partial \tilde{g}^{**}}{\partial \boldsymbol{\Sigma}} = \frac{\mathrm{tr}\,\partial\,(n\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{V}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}}. \tag{4.23}$$

In (4.23), we set $\partial \tilde{g}^{**}/\partial \boldsymbol{\Sigma} = \mathbf{0}$. Then, $n\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{V}\boldsymbol{\Sigma}^{-1} = \mathbf{0}$, and the estimate of $\boldsymbol{\Sigma} = n^{-1}\mathbf{V}$. Thus the error covariance matrix, $\boldsymbol{\Sigma}$, can be replaced by its maximum likelihood estimate, $n^{-1}\hat{\mathbf{V}}$, where $\hat{\mathbf{V}} = (\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\hat{\mathbf{B}})'(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\hat{\mathbf{B}})$.

We alternate the two steps until convergence of (4.17) is reached. When the errors are assumed to be multivariate normal and the error covariance matrix is estimated by $n^{-1}\hat{\mathbf{V}}$, the estimators obtained by maximizing (4.17) are efficient (e.g., Velu, 1991). As shown above, maximizing (4.17) comes down to minimizing $f(\mathbf{B}, \boldsymbol{\Sigma}) = \frac{1}{2}\mathrm{tr}[(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}\mathbf{B})']$ with respect to $\mathbf{B}$ and $\boldsymbol{\Sigma}$. The parameter estimates obtained by minimizing this generalized least squares criterion are asymptotically equivalent to the maximum likelihood estimates (e.g., Bradley, 1973; Goldstein, 1986).

Besides providing the efficient estimator, the additional assumption of normality also enables us to perform significance tests without recourse to resampling methods such as the bootstrap method. In general, those significance tests can be thought of as a comparison of goodness of fit between two nested models. Two models are called nested if one is a restrictive version of the other. There are three asymptotically equivalent tests of differences between two nested models: the likelihood ratio (LR), the Wald test, and the Lagrangian multiplier (LM) test (e.g., Buse, 1985; Engle, 1984; Lee & Bentler, 1980; Satorra, 1989; Silvey, 1959). The LR test compares the value of

the likelihood function for a less restrictive model with that of a more restrictive model. Thus, it requires fitting of both restrictive and less restrictive models. On the other hand, the Wald and LM tests do not require fitting both models. For the Wald test, only the less restrictive model is fitted, whereas for the LM test, only the more restrictive model is fitted.

These tests are employed to compare the fit of more restrictive to less restrictive models. As a special case, a specific model of interest can be compared against a saturated model, which provides the goodness of fit of the specific model relative to the saturated model. As another example, a model with additional constraints can be tested against a model without the constraints, which allows us to investigate the validity of the constraints. When a more restrictive model (or specified model) is rejected, we may try to modify the model, removing constraints on parameters. For this purpose, we can use the modification index (Jöreskog & Sörbom, 1984), which is analogous to the univariate LM test. The modification index tells us which constrained parameter should be freed from zero to obtain the maximum improvement in fit, without fitting models that eliminate one constraint at a time.

For significance tests in extended redundancy analysis, we employ the LM test. The LM test statistic is given by

$$\mathrm{LM} = \mathbf{v}(\boldsymbol{\theta}_r)' \mathbf{J}^{-1}(\boldsymbol{\theta}_r)\mathbf{v}(\boldsymbol{\theta}_r), \tag{4.24}$$

where $\boldsymbol{\theta}_r$ denotes the vector of parameters in a more restrictive model, $\mathbf{v}(\boldsymbol{\theta}_r)$ is the vector of the first-order derivative of the log likelihood function for a less restrictive

model evaluated at $\boldsymbol{\theta}_r$, called the efficient score, and $\mathbf{J}(\boldsymbol{\theta}_r)$ is the matrix of the expected second-order derivatives of the log likelihood function for the less restrictive model evaluated at $\boldsymbol{\theta}_r$, called the information matrix.

As an example, suppose that $\boldsymbol{\theta}$ is the vector consisting of parameters without any additional constraints imposed, i.e., $\boldsymbol{\theta} = [\text{vec}(\mathbf{W})', \text{vec}(\mathbf{A}')']'$. When we set $\boldsymbol{\Sigma} = n^{-1}\hat{\mathbf{V}}$, then, maximizing (4.17) reduces to minimizing

$$\begin{aligned} f(\boldsymbol{\theta}) &= \frac{1}{2}\text{tr}[(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')'(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')\boldsymbol{\Sigma}^{-1}] \\ &= \frac{1}{2}[\text{vec}(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')]'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})[\text{vec}(\mathbf{Z}^{(1)}-\mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')] \\ &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Pi}\boldsymbol{\mu}, \end{aligned} \tag{4.25}$$

where $\boldsymbol{\mu}= \text{vec}(\mathbf{Z}^{(1)}- \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')$, and $\boldsymbol{\Pi} = \boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}$. We note that

$$\begin{aligned} \boldsymbol{\mu} &= \text{vec}(\mathbf{Z}^{(1)}) - (\mathbf{I}\otimes\mathbf{Z}^{(2)})\text{vec}(\mathbf{W}\mathbf{A}') \\ &= \text{vec}(\mathbf{Z}^{(1)}) - (\mathbf{I}\otimes\mathbf{Z}^{(2)})(\mathbf{A}\otimes\mathbf{I})\text{vec}(\mathbf{W}) \\ &= \text{vec}(\mathbf{Z}^{(1)}) - (\mathbf{I}\otimes\mathbf{Z}^{(2)})(\mathbf{I}\otimes\mathbf{W})\text{vec}(\mathbf{A}'). \end{aligned}$$

Then, $\mathbf{v}(\boldsymbol{\theta})$ can be expressed as

$$\mathbf{v}(\boldsymbol{\theta}) \equiv \frac{1}{2}\frac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{1}{2}\frac{\partial\boldsymbol{\mu}'}{\partial\boldsymbol{\theta}}\boldsymbol{\Pi}\boldsymbol{\mu} = -\begin{bmatrix} \mathbf{A}'\otimes\mathbf{I} \\ \mathbf{I}\otimes\mathbf{W}' \end{bmatrix}(\mathbf{I}\otimes\mathbf{Z}'^{(2)})\boldsymbol{\Pi}\boldsymbol{\mu}. \tag{4.26}$$

Also, $\mathbf{J}(\boldsymbol{\theta})$ can be written as

$$\mathbf{J}(\boldsymbol{\theta}) \equiv -E\left[\frac{\partial^2 f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \left[\left(\frac{\partial\boldsymbol{\mu}'}{\partial\boldsymbol{\theta}}\right)\boldsymbol{\Pi}\left(\frac{\partial\boldsymbol{\mu}'}{\partial\boldsymbol{\theta}}\right)'\right]. \tag{4.27}$$

These $\mathbf{v}(\boldsymbol{\theta})$ and $\mathbf{J}(\boldsymbol{\theta})$ are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_r$. The LM asymptotically follows a chi-square distribution with degrees of freedom equal to the difference in degrees of freedom between restrictive and less restrictive models under the hypothesis that the more restrictive model is true. When a specific model is compared against the saturated model, the asymptotic distribution of LM is chi-square with degrees of freedom equal to the difference between the number of the parameters in the saturated model and the specific model under the hypothesis that the model to be evaluated is true.

van der Leeden (1990) pointed out that it would be difficult to specify the saturated model in redundancy analysis, since the saturated model would arise if each individual observation had its own set of parameters. This is also the case in our method. The problem of formulating a suitable saturated model interferes with testing the goodness of fit of a specific model. However, the multivariate regression model without any rank constraints on the matrix of regression coefficients (the so-called full-rank model) may be used as the most general model (van der Leeden, 1990). What we can do with confidence is to test a more restrictive model against a less restrictive model from a set of nested models. Nested models may be created by successively reducing the ranks of $\mathbf{B} = \mathbf{W}\mathbf{A}'$(e.g., Velu, 1991). It may also be of interest to compare a specific model against the null model in which it is hypothesized that $\mathbf{B} = \mathbf{O}$ (or rank($\mathbf{B}$) = 0). In this case the former is the less restrictive model.

The univariate LM test, which is equivalent to the modification index, is given by

$$\left[\frac{\partial f(\boldsymbol{\theta})}{\partial \theta_i} \right]^2_{\theta_i = (\theta_r)_i} \mathbf{J}^{-1}(\boldsymbol{\theta}_r)_{ii}, \tag{4.28}$$

where $(\boldsymbol{\theta}_r)_i$ indicates the ith element of $\boldsymbol{\theta}_r$, and $\mathbf{J}^{-1}(\boldsymbol{\theta}_r)_{ii}$ represents the ith diagonal element of $\mathbf{J}^{-1}(\boldsymbol{\theta}_r)$ (e.g., Bollen, 1989, pp. 298-299). The only difference between a less restrictive model and a more restrictive model is in the constraints on θ_i. The univariate LM test is an asymptotic chi-square test with one degree of freedom to assess the improvement in fit when θ_i is freed.

In conclusion, normality is not necessary for our method. If it is assumed, nonetheless, our method can provide efficient estimators and also enables us to perform statistical inferences such as the tests of hypotheses without recourse to resampling methods such as the bootstrap method.

4. 6. Examples

In this section, we illustrate only data transformations, robust estimation, and efficient estimation, since missing data can be easily handled in the process of data transformations. Robust estimation and efficient estimation procedures involve specifying an iteratively updated row-side metric matrix and column-side metric matrix, respectively. Throughout this section, we fit the two latent variable model for the basic health indicator data, given in Section 2.5.

To exemplify data transformations, two observed endogenous variables, that is, IMR and MMR, were monotonically transformed, based on the procedure described in Section 4.1. We applied Kruskal's (1964) primary least squares monotonic transformation to those variables. This indicated that observation categories were order-preserved but tied observations might become untied. The least squares monotonic transformations of the variables are shown in Figures 4.1a and 4.1b.

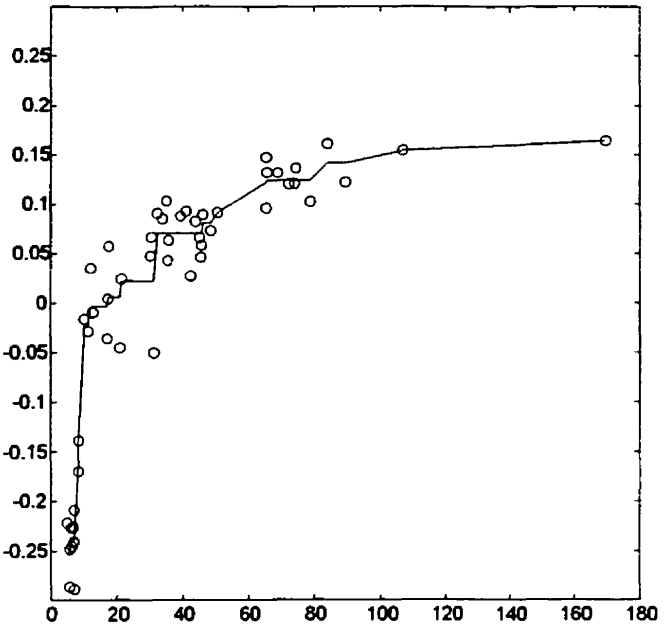

Figure 4.1a. The least squares monotonic transformation of variable IMR.

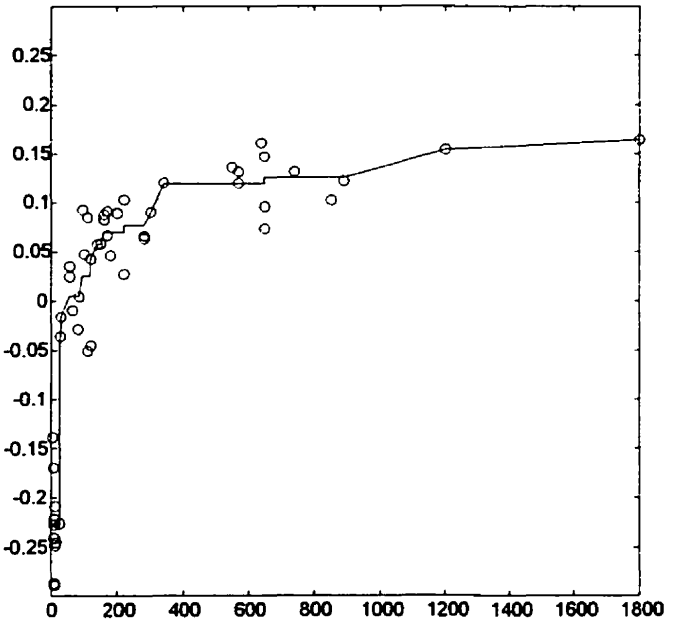

Figure 4.1b. The least squares monotonic transformation of variable MMR.

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 (.9571), while providing similar interpretations of parameter estimates as those obtained when the variables were treated as numerical. The correlation between two latent variables was, however, counter-intuitively low (-.04), compared with that from the untransformed case. This leads some difficulty in interpretations of the obtained solutions.

We then applied Ramsay's (1998) smooth monotonic transformation to the same endogenous variables. (The matlab codes for the smooth monotonic transformation were kindly provided by Jim Ramsay.) The estimated smooth monotonic transformations of the same variables are provided in Figures 4.2a and 4.2b. For each of the endogenous variables, we used low dimensional B-spline base functions with a single interior knot, which was positioned at the median, and chose the value of the smoothing parameter as 0.01.

The model fit was .8589. This is worse than that from the least squares monotonic transformation. It is because the smooth monotone transformation is more restrictive, removing more wiggles of the transformed scores. Nevertheless, the smooth monotonic transformation showed a much better model fit than that obtained from the non-transformed case. More significant was that Ramsay's smooth monotonic transformation also provided almost the same correlation between two latent variables (.46) as that obtained from the non-transformed case (.47).

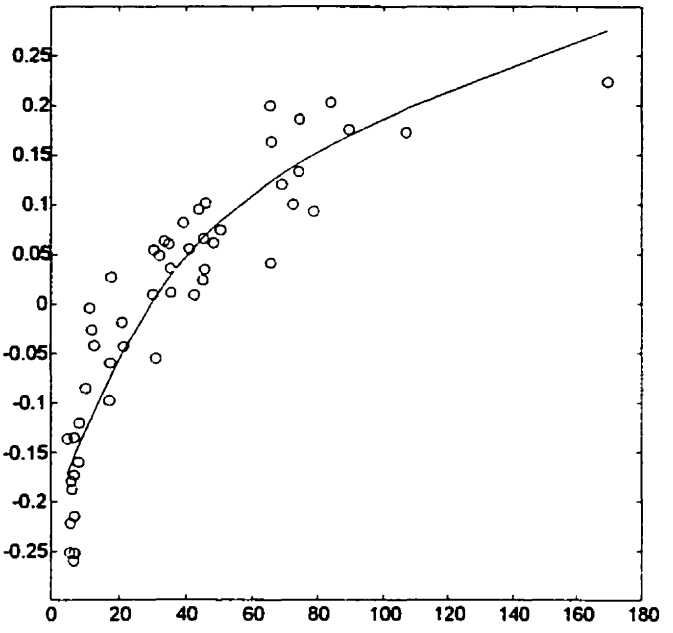

Figure 4.2a. The smooth monotonic transformation of variable IMR.

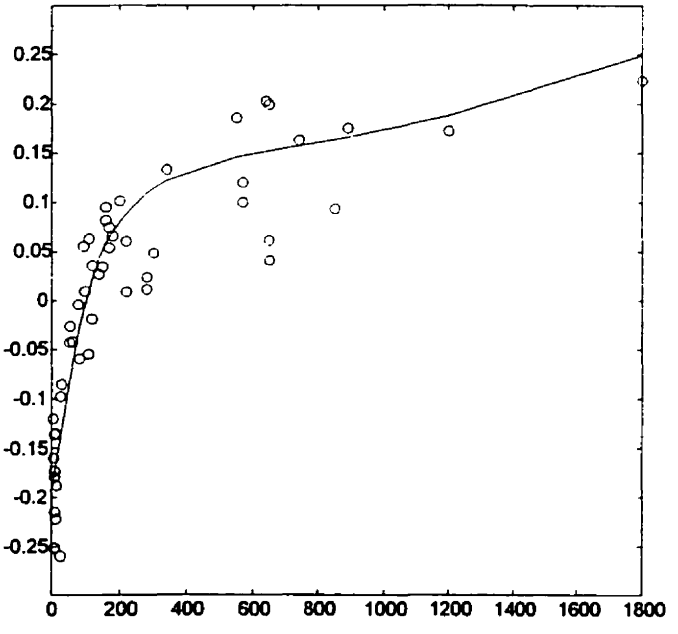

Figure 4.2b. The smooth monotonic transformation of variable MMR.

This seems to make the obtained solutions more interpretable. As the smooth monotonic transformation is applied to endogenous variables, however, the conventional distributional assumptions of regression analysis (i.e., errors are assumed to be normal, and independently and identically distributed with mean 0 and variance σ^2) are needed due to the employment of the maximum likelihood estimation.

To illustrate robust estimation, the IRLS procedure described in Section 4.4 was implemented, and the robust extended redundancy analysis was applied to fit the two latent variable model. The variable sensitivity factor was chosen to be 6. Results of fitting the model by the robust estimation method are presented in Figure 4.3. The bootstrapped standard errors and the critical ratios obtained with 100 bootstrap samples are given in Table 4.1.

Due to the effect of robust estimation, the model fit was equal to .7379. It turned out to be significant in terms of its critical ratio (12.0). The squared multiple correlations of IMR and MMR were .79 and .66, respectively. They also turned out to be significant in terms of their critical ratios (18. 1 and 7.8 for IMR and MMR, respectively). Interpretations of the solutions obtained from the robust method were essentially the same as those from ordinary extended redundancy analysis, given in Section 2.5. In the robust estimation case, however, GDP turned out to be more strongly associated with SE than FEDU, which was opposite to the case of ordinary extended redundancy analysis. This may indicate that some countries had exceptionally high or low GDP relative to the other countries, acting as outliers, and the adjustment of them led to a better association with SE.

Fit = .7379

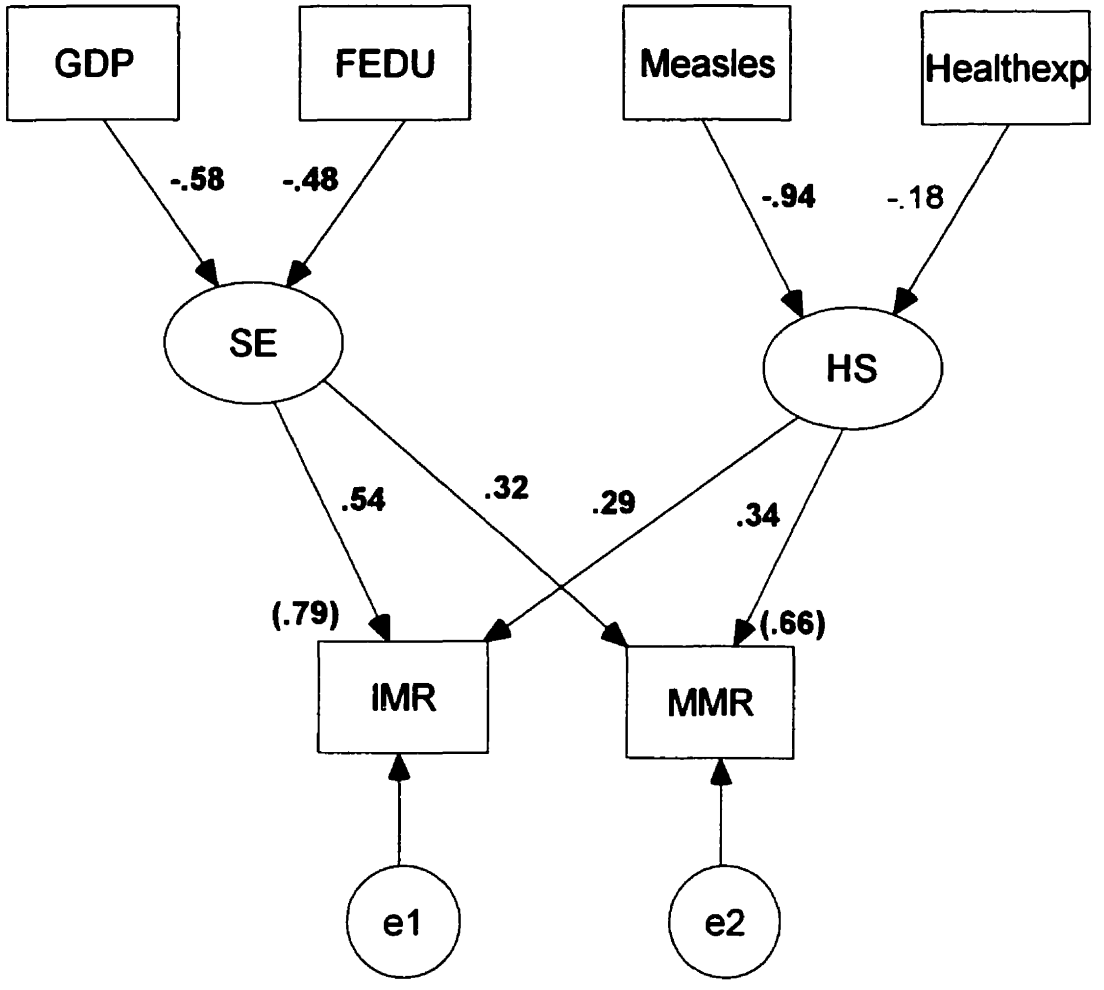

Figure 4.3. The two latent variable model for the WHO data (robust estimation).

Table 4.1. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from robust estimation of the two latent variable model for the WHO data.

	Estimate	SE	CR
w_1	-.58	.18	-3.2
w_2	-.48	.18	-2.7
w_3	-.94	.17	-5.5
w_4	-.18	.26	-0.7
a_1	.54	.10	5.4
a_2	.32	.10	3.2
a_3	.29	.10	2.9
a_4	.34	.13	2.6

The correlation between the latent variables (.47) turned out significant (the bootstrapped critical ratio = 3.1).

To demonstrate efficient estimation under the normality assumption, we compared the two latent variable model against the full-rank model (i.e., the regular multivariate regression analysis model). The efficient parameter estimates of the two latent variable model are presented in Figure 4.4. Their standard errors (i.e., the squared diagonal elements of $\mathbf{J}^{-1}(\boldsymbol{\theta}_r)$) and critical ratios are given in Table 4.2. The fit of the two latent variable model was quite good (LM = .30 with df = 2), indicating that the two latent variable model was more appropriate than the full-rank model. (The degrees of freedom of the full-rank model were equal to $n - 8$, while those of the two latent variable model were equal to $n - 6$ due to the identification restriction, $\text{diag}(\mathbf{W}'\mathbf{Z}^{(2)\prime}\mathbf{Z}^{(2)}\mathbf{W}) = \mathbf{I}_2$, where $\mathbf{I}_2$ is an identity matrix of order 2.) The squared multiple correlations of IMR and MMR were .64 and .57, respectively. Less variance of IMR was accounted for by SE than in the least squares estimation case (.73), given in Figure 2.5, whereas almost the same variance of MMR was explained by HS. The correlation between the latent variables was equal to .44, which was about the same as that from the least squares estimation (.46). Variable GDP was found to be a bit more strongly associated with SE (.54) than FEUD (.52), which was opposite to the least squares estimation case. Nonetheless, interpretations of the efficient estimates were essentially the same as those of the least squares counterparts.

Furthermore, we compared a more restrictive model, in which the component weight for Healthexp (i.e., w_4) was additionally constrained to be equal to zero, against

LM = .30
df = 2

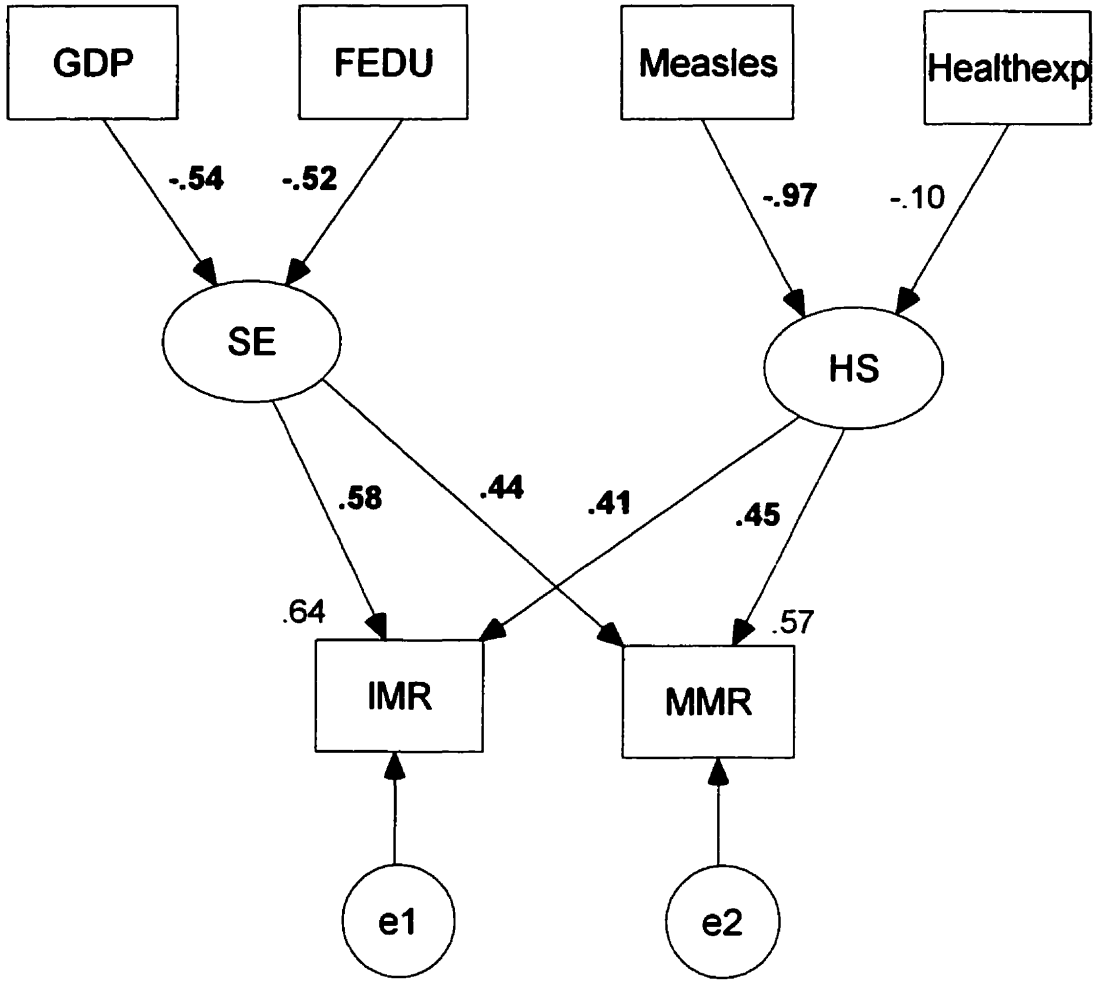

Figure 4.4. The two latent variable model for the WHO data (efficient estimation)

Table 4.2. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from efficient estimation of the two latent variable model for the WHO data.

	Estimate	SE	CR
w_1	-.54	.12	-4.5
w_2	-.52	.13	-4.0
w_3	-.97	.08	-12.1
w_4	-.10	.09	-1.1
a_1	.58	.15	3.7
a_2	.44	.10	4.4
a_3	.41	.16	2.6
a_4	.45	.11	4.1

the two latent variable model. The efficient parameter estimates of the constrained two latent variable model are provided in Figure 4.5. Their standard errors and critical ratios are given in Table 4.3. In Chapter 2 the same comparison was carried out on the basis of the least squares estimation, depicted in Figure 2.6.

The constrained model showed a fairly good fit (LM = .21 with df = 1), indicating that it is more appropriate than the unconstrained two latent variable model. This is consistent with the conclusion from the least squares estimation. The squared multiple correlations of IMR and MMR were equal to .64 and .57, respectively. The correlation between the latent variables was equal to .40. The efficient estimates of the constrained model could be similarly interpreted to the least squares counterparts.

The LM test between the two latent variable model and the constrained two latent variable model was equivalent to the univariate LM test, since the only difference between them was in the zero constraint on w_4. Thus, the modification index of the constrained parameter was identical to the value of the LM test (.21). This indicates that removing the constraint from the constrained model does not improve the model fit significantly.

LM = .21
df = 1

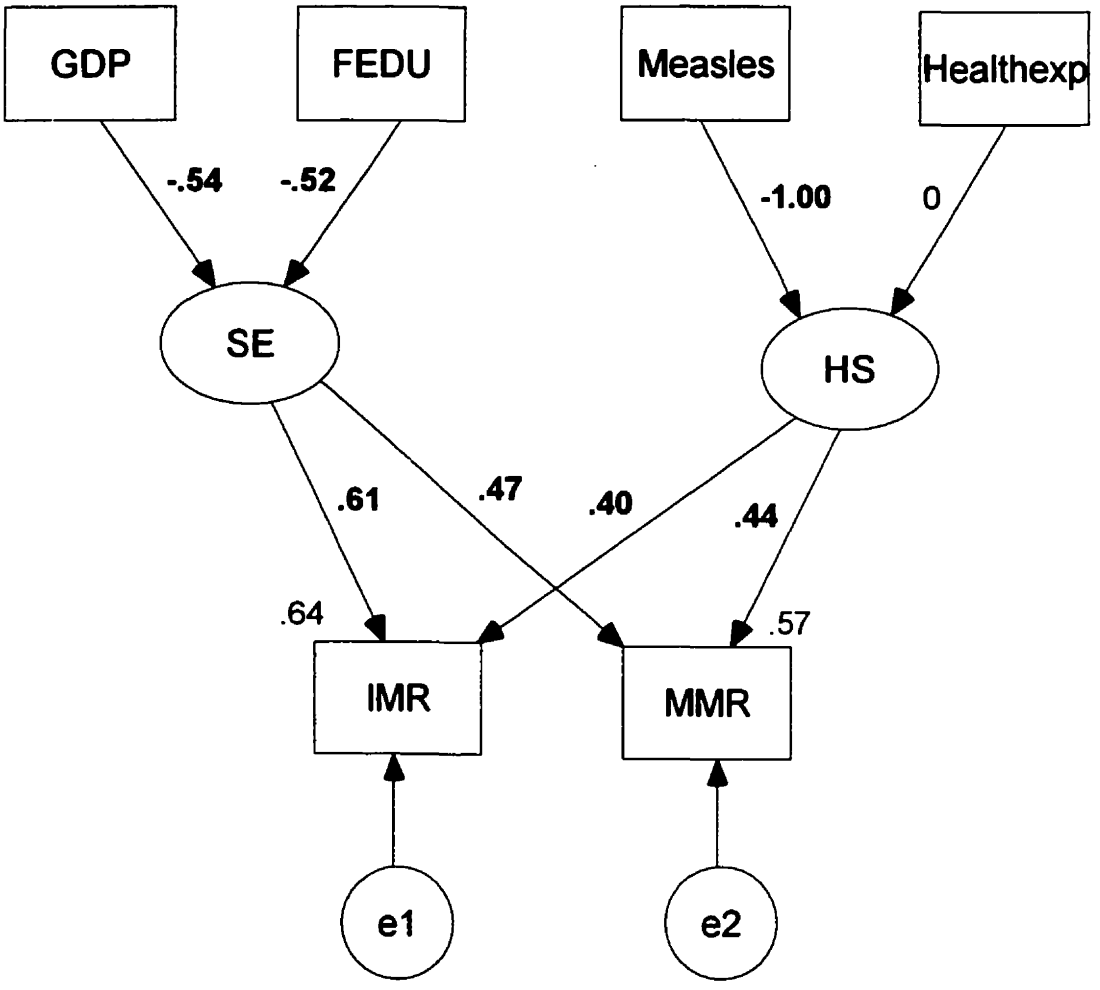

Figure 4.5. The constrained two latent variable model for the WHO data (efficient estimation)

Table 4.3. The parameter estimates, and their standard errors (SE) and critical ratios (CR) obtained from efficient estimation of the constrained two latent variable model for the WHO data.

	Estimate	SE	CR
w_1	-.54	.16	-3.4
w_2	-.52	.17	-3.1
a_1	.61	.04	15.3
a_2	.47	.06	7.8
a_3	.40	.07	5.7
a_4	.44	.08	5.5

Chapter 5

Discussion

A new method, called extended redundancy analysis, is proposed for analysis of structural equation models, which is viewed as an extension of redundancy analysis. One of the major characteristics of the method is to estimate latent variables as exact linear combinations of observed exogenous variables. This enables us to avoid improper solutions which frequently occur in covariance structure analysis. Another crucial feature is to employ a well-defined least squares criterion to estimate model parameters. This allows for an overall model fit, and also ensures an optimality of obtained solutions, which is not guaranteed in PLS. An alternating least squares algorithm is developed to optimize the criterion. Our experience is that the algorithm is quite efficient. It converges fast, and seems to be hardly afflicted by the non-global minimum problem.

The proposed method is simple yet versatile enough to fit various complex relationships among variables, including direct effects of observed variables and higher-order latent variables. Moreover, it is able to perform multi-sample comparisons. The method can deal with data transformations, missing data, and robust estimation in a simple way, which further broadens its capacity. When the normality assumption is satisfied, the present method can also allow for tests of statistical significance.

Some researchers claim that linear components should not be called latent variables (e.g., MacCallum & Browne, 1993; McDonald, 1996). They distinguish a latent variable as an unobserved, error-free, and more generalizable variable that accounts for observed variables with sampling error added, usually represented by common factors (Velicer & Jackson, 1990). On the other hand, latent variables estimated as linear combinations of observed variables are obtained without taking account of measurement errors in observed variables. In practical sense, however, it is difficult to find any substantial differences between factor score estimates and component scores (Velicer, 1976; Velicer & Fava, 1987). More importantly, component loadings can be used as reasonable and interpretable estimates when factor loadings cannot (Kiers *et al.*, 1996). Furthermore, if a population distribution is known (e.g., a normal distribution), the proposed method can provide efficient estimates. Further studies are, however, needed to compare the quality of parameter estimates obtained from our method with those from covariance structure analysis and PLS.

A more fundamental limitation of the present method is that it cannot accommodate as various relationships among variables as covariance structure analysis or PLS can do. One example is that our method can only accommodate formative relationships between latent variables and observed exogenous variables so far. That is, the observed exogenous variables are always assumed to be multiple causes of a latent variable. However, our method may capture reflective relationships among latent and the observed exogenous variables (i.e., the observed exogenous variables are affected by an underlying construct) by exploiting the principle of principal covariates

regression (de Jong & Kiers, 1992).

In principal covariates regression (PCR), it is assumed

$$\begin{aligned} \mathbf{Z}_1 &= \mathbf{F}\mathbf{A}' + \mathbf{E}, \\ \mathbf{Z}_2 &= \mathbf{F}\mathbf{D}' + \mathbf{E}^*, \\ \mathbf{F} &= \mathbf{Z}_2\mathbf{W}, \end{aligned} \tag{5.1}$$

where $\mathbf{D}'$ is a loading matrix, and $\mathbf{E}^*$ is the matrix of measurement errors of $\mathbf{Z}_2$. The other matrices are defined as in (2.1). In PCR, latent variables are determined in such a way that they account for some of the variance of $\mathbf{Z}_2$ while they also predict $\mathbf{Z}_1$. The fitting criterion in PCR is given by

$$\begin{aligned} f &= \mathrm{SS}([\mathbf{Z}_1 \vdots \mathbf{Z}_2] - \mathbf{Z}_2\mathbf{W}[\mathbf{A}' \vdots \mathbf{D}']) \\ &= \mathrm{SS}(\mathbf{Z} - \mathbf{Z}_2\mathbf{W}\mathbf{T}), \end{aligned} \tag{5.2}$$

subject to $\mathbf{W}'\mathbf{Z}_2'\mathbf{Z}_2'\mathbf{W}' = \mathbf{I}$, where $\mathbf{Z} = [\mathbf{Z}_1 \vdots \mathbf{Z}_2]$, and $\mathbf{T} = [\mathbf{A}' \vdots \mathbf{D}']$. Criterion (5.2) is essentially the same as (2.8). Optimizing this least squares criterion amounts to calculating the eigenvalue decomposition of $[\mathbf{Z}_1\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{Z}_2']$ (see de Jong & Kiers, 1992, for detailed derivations).

PCR may be extended to more than two sets of data. The extended PCR model may be written as follows.

$$\begin{aligned} \mathbf{Z}^{(1)} &= \mathbf{F}\mathbf{A}' + \mathbf{E}, \\ \mathbf{Z}^{(2)} &= \mathbf{F}\mathbf{D}' + \mathbf{E}^*, \\ \mathbf{F} &= \mathbf{Z}^{(2)}\mathbf{W}, \end{aligned} \tag{5.3}$$

where $\mathbf{D}'$ is a matrix of loadings, and $\mathbf{E}^*$ contains measurement error of $\mathbf{Z}^{(2)}$. The other matrices are defined as in (2.6). To estimate parameters, we may aim to minimize the

following criterion

$$f = \mathrm{SS}([\mathbf{Z}^{(1)} \vdots \mathbf{Z}^{(2)}] - \mathbf{Z}^{(2)}\mathbf{W}[\mathbf{A}' \vdots \mathbf{D}'])$$
$$= \mathrm{SS}(\mathbf{Z} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{T}), \qquad (5.4)$$

where $\mathbf{Z} = [\mathbf{Z}^{(1)} \vdots \mathbf{Z}^{(2)}]$, and $\mathbf{T} = [\mathbf{A}' \vdots \mathbf{D}']$. In (5.4), $\mathbf{W}$ and $\mathbf{T}$ can be structured according to the model to be fitted. Criterion (5.4) is essentially of the same form as (2.8), and the same ALS algorithm can be used to minimize (5.4).

By incorporating the PCR feature, our method can deal with both reflective and formative relationships among latent and observed exogenous variables. This is comparable to PLS, which is also capable of handling both kinds of relationships. In covariance structure analysis, on the other hand, the reflective relationships are typically assumed because they are consistent with its statistical algorithm based on all the covariances among observed variables (Chin, 1998). Modeling formative relationships in covariance structure analysis requires certain conditions (e.g., a latent variable defined by linear combinations of observed variables needs to emit at least two paths to different latent variables) in order to avoid some identification problems (MacCallum & Browne, 1993). Yet it may be often difficult to satisfy those conditions in model specification.

Another noticeable drawback of the present method is that it is impossible to assume any latent variables for the observed endogenous variables. This happens because our method is an extension of redundancy analysis, whose main goal is to obtain linear components of the observed exogenous variables that explain the maximum variance of the observed endogenous variables, so that no linear components

are considered for the observed endogenous variables. Hence, this problem seems inevitable unless we propose models outside the realm of redundancy analysis. We may consider canonical regression analysis (e.g., van der Leeden, 1990, p. 47) as a potential candidate to resolve the problem.

The model for canonical regression analysis may be written as

$$\begin{aligned} \mathbf{Z}_1\mathbf{N} &= \mathbf{Z}_2\mathbf{W}\mathbf{A}' + \mathbf{E}, \\ \mathbf{F}_1 &= \mathbf{F}_2\mathbf{A}' + \mathbf{E}, \end{aligned} \tag{5.5}$$

where $\mathbf{N}$ is a matrix of component weights for $\mathbf{Z}_1$, $\mathbf{F}_1 = \mathbf{Z}_1\mathbf{N}$, $\mathbf{F}_2 = \mathbf{Z}_2\mathbf{W}$, and the other parameter matrices are defined as in (2.1). From (5.5), we see that the canonical regression analysis model specifies an asymmetric relationship between two sets of latent variables, one from a set of the observed exogenous variables, and the other from a set of the observed endogenous variables. This is different from canonical correlation analysis that is concerned with a symmetric relationship between the two sets of linear components. To estimate model parameters, we may minimize the criterion as follows:

$$\begin{aligned} f &= \mathrm{SS}(\mathbf{Z}_1\mathbf{N} - \mathbf{Z}_2\mathbf{W}\mathbf{A}') \\ &= \mathrm{SS}(\mathbf{F}_1 - \mathbf{F}_2\mathbf{A}'), \end{aligned} \tag{5.6}$$

with respect to $\mathbf{N}$, $\mathbf{W}$, and $\mathbf{A}'$.

Canonical regression analysis may be extended in a similar way to extended redundancy analysis. This extension may be called extended canonical regression analysis (ECRA). The ECRA model may be written as

$$\mathbf{Z}^{(1)}\mathbf{N} = \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}' + \mathbf{E},$$
$$\mathbf{F}^{(1)} = \mathbf{F}^{(2)}\mathbf{A}' + \mathbf{E}, \tag{5.7}$$

where $\mathbf{N}$ is a matrix of component weights for $\mathbf{Z}^{(1)}$, $\mathbf{F}^{(1)} = \mathbf{Z}^{(1)}\mathbf{N}$, and $\mathbf{F}^{(2)} = \mathbf{Z}^{(2)}\mathbf{W}$, and the other matrices are analogously defined as those in (2.6). Matrices $\mathbf{N}$, $\mathbf{W}$, and $\mathbf{A}'$ can be structured according to the model to be fitted. The fitting criterion may be given as

$$f = \mathrm{SS}(\mathbf{Z}^{(1)}\mathbf{N} - \mathbf{Z}^{(2)}\mathbf{W}\mathbf{A}')$$
$$= \mathrm{SS}(\mathbf{F}^{(1)} - \mathbf{F}^{(2)}\mathbf{A}'). \tag{5.8}$$

We may use an ALS algorithm to optimize the criterion, which updates alternately each of the parameter matrices (i.e., $\mathbf{N}$, $\mathbf{W}$, and $\mathbf{A}'$) with the others fixed.

It seems promising to incorporate the PCR feature into ECRA. This may provide a comprehensive component-based analysis for structural equation models. Future research is needed to investigate the feasibility of this approach.

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.

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

Beaton, A. E. & Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147-185.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying influential Data and Sources of Collinearity*. New York: Wiley.

Bock, R. D., & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika*, **31**, 507-534.

Böckenholt, U., & Böckenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, **55**, 633-639.

Böckenholt, U., & Takane, Y. (1994). Linear constraints in correspondence analysis. In M. J. Greenacre & J. Blasius (Eds.), *Correspondence Analysis in Social Sciences* (pp. 112-127). London: Academic Press.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, **50**, 229-242.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of Royal Statistical Society Series B*, **26**, 211-252.

Bradley, E. L. (1973). The equivalence of maximum likelihood and weighted least squares estimates in the exponential family. *Journal of American Statistical Association*, **68**, 199-200.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis* (pp. 72-141). Cambridge: Cambridge University Press.

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

Buse, A. (1985). The likelihood ratio, Wald, and Langrange multiplier tests: An expository note. *The American Statistician*, **36**, 153-157.

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *Management Information Systems Quarterly*, **22**, 7-16.

Cook, R. D., & Weisberg, S. (1982). Residuals and Influence in Regression. New York: Chapman and Hall.

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.

Davies, P. T., & Tso, M. K.-S. (1982). Procedures for reduced-rank regression. *Applied Statistics*, **31**, 244-255.

de Jong, S., & Kiers, H. A. L. (1992). Principal covariates regression.

Chemometrics and Intelligent Laboratory Systems, **14**, 155-164.

de Leeuw, J. (1989). Fitting reduced rank regression models by alternating maximum likelihood. UCLA Statistics Series 35. Department of Statistics, University of California at Los Angeles.

de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 471-503.

Engle, R. F. (1984). Wald, likelihood ratio, and Langrange multiplier tests in econometrics. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics Vol II* (pp. 775-826). Amsterdam: North Holland.

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

Efron, B., & Tibshirani, R. J. (1998). *An Introduction to the Bootstrap*. CRC Press. Boca Raton, Florida.

Escoufier, Y. (1987). The duality diagram: A means for better practical applications. In P. Legendre & L. Legendre (Eds.), *Development in numerical ecology* (pp. 139-156). Berlin: Springer.

Falk, R. F. & Miller, N. (1992). *A Primer for Soft Modeling*. Akron, Ohio: The University of Akron Press.

Fornell, C., Barclay, D. W., & Rhee, B. (1988). A model and simple iterative algorithm for redundancy analysis. *Multivariate Behavioral Research*, **23**, 349-360.

Fornell, C., & Bookstein, F. (1982). Two structural equation models: LISREL

and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, **19**, 440-452.

Fornell, C., & Cha, J. (1994). Partial least squares, In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 52-78). Oxford: Blackwell Publisher.

Gabriel, K. R., & Zamir, S. (1979). Low rank approximation of matrices by least squares with any choice of weights. *Technometrics*, **21**, 489-498.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.

Goldstein, H. (1986). Multilevel mixed linear model analysis with iterative generalized least squares. *Biometrika*, **73**, 43-56.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.

Griep, M. I., Wakeling, I. N., Vankeerberghen, P., & Massart, D. L. (1995). Comparison of semirobust and robust partial least squares procedures. *Chemometrics and Intelligent Laboratory Systems*, **29**, 37-50.

Guttman, L. (1953). Image theory for the structure of quantitative variables. *Psychometrika*, **9**, 277-296.

Hayduk, L. A. (1987). *Structural Equation Modeling with LISREL*. Baltimore: Johns Hopkins University Press.

Hoyle, R. H. (1995). *Structural Equation Modeling, Concepts, Issues, and Applications*. Thousand Oaks: Sage Publications.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, **51**, 563-566.

Israëls, A. Z. (1984). Redundancy analysis for qualitative variables. *Psychometrika*, **49**, 331-346.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, **5**, 248-264.

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

Jöreskog, K. G. (1973). A generating method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences* (pp. 85-112). New York: Academic Press.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: specification, estimation, and testing. In P. R. Krishnaiah (Ed.), *Applications of Statistics* (pp. 265-287). Amsterdam: North Holland.

Jöreskog, K. G. (1993). Testing structural equation models, In K. A. Bollen & S. J. Long (Eds.) *Testing Structural Equation Models* (pp. 294-316). London: Sage publications.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **10**, 631-639.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Method*. Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and*

applications. Chicago: Scientific Software.

Jöreskog, K. G. & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: Historical and comparative Aspects. In H. Wold and K. G. Jöreskog (Eds.), *Systems Under Indirect Observation*: *Causality, Structure, Prediction, Part I* (pp. 263-270). Amsterdam: North-Holland.

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.

Kiers, H. A. L., Takane, Y., & ten Berge, J. M. F. (1996). The analysis of multitrait-multimethod matrices via constrained components analysis. *Psychometrika*, **61**, 601-628.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Publications.

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

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1988). Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset associations. *Psychological Bulletin*, **104**, 282-289.

Lee, S.-Y., & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal*, **14**, 121-136.

Loehlin, J. C. (1998). *Latent Variable Models*: *An Introduction to Factor, Path,*

and Structural Analysis. Mahwah: Lawrence Erlbaum Associates.

Lohmöller, J. B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. New York: Springer-Verlag.

Lyttkens, E. (1968). On the fixed-point property of Wold's iterative estimation method for principal components. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 335-350). New York: Academic Press.

Lyttkens, E. (1973). The fixed-point method for estimating interdependent systems with the underlying model specification, *Journal of the Royal Statistical Society*, **A 136**, 353-394.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, **114**, 533-541.

McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, **31**, 239-270.

McDonald, R. P., & Mulaik, S. A. (1979). Determinacy of common factors: A nontechnical review. *Psychological Bulletin*, **86**, 297-308.

Meijer, E. (1998). *Structural Equation Models for Nonnormal Data*. Leiden: DSWO Press.

Meredith, W., & Millsap, R. E. (1985). On component analysis. *Psychometrika*, **50**, 495-507.

Meulman, J. (1986). *A Distance Approach to Nonlinear Multivariate Analysis*. Leiden: DSWO Press.

Micceri, T. (1989). The unicorn, the normal curve, and other improvable creatures. *Psychological Bulletin*, **105**, 156-166.

Millsap, R. E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika*, **53**, 123-134.

Muthén, B. O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, **43**, 551-560.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, **49**, 115-132.

Muthén, B. O. (1987). *LISCOMP. Analysis of Linear Structural Equations With A Comprehensive Measurement Model. User's Guide*. Mooresville: Scientific Software.

Muthén, B. O., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, **45**, 19-30.

Park, Y. (1996). Das soziale Netzwerk in der fruhen Adoleszens: eine vergleichende Untersuchung zweier Generationen aus Korea und Deutchland. Unpublished doctoral dissertation. Koln: University of Koln.

Poffhoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313-326.

Ramsay, J. O. (1988). Monotone regression splines in action (with discussion). *Statistical Science*, **3**, 425-461.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of Royal*

Statistical Society B, **60**, 365-375.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyã* A, **26**, 329-358.

Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 3-22). Amsterdam: North Holland.

Reinsel, G. C., & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression*. New York: Springer-Verlag.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, **54**, 131-151.

Schönemann, P. H., & Steiger, J. H. (1976). Regression component analysis. *British Journal of Mathematical and Statistical Psychology*, **29**, 175-189.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. New York: Wiley.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Silvey, S. D. (1959). The Langrangian multiplier test. *Annals of Mathematical Statistics*, **30**, 389-407.

Singh, A. (1996). Outliers and robust procedures in some chemometric applications. *Chemometrics and Intelligent Laboratory Systems*, **33**, 75-100.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's: Some interesting parallels. *Psychometrika*, **44**, 157-166.

Takane, Y. (1978). A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent - evaluations. *Japanese Psychological Research*, 20, 105-114.

Takane, Y., & Hunter, M. A. (2000). Constrained principal component analysis: A comprehensive theory. 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.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, **56**, 97-120.

Takane, Y., Yanai, H., & Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, **56**, 667-684.

ten Berge, J. M. F. (1993). Least squares optimization in multivariate analysis. Leiden: DSWO Press.

ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvalue technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167-1179.

Tso, M. K.-S. (1981). Reduced-rank regression and canonical analysis. *Journal of Royal Statistical Society B*, **43**, 183-189.

van Buuren, S. (1990). *Optimal Scaling of Time Series*. Leiden: DSWO Press.

van der Burg, E. (1988). *Nonlinear Canonical Correlation and Some Related Techniques*. Leiden: DSWO Press.

van der Leeden, R. (1990). *Reduced Rank Regression with Structured Residuals*. Leiden: DSWO Press.

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

Velicer, W. F. (1976). The relation between factor score estimates, image scores, and principal component scores. *Educational and Psychological Measurement*, **36**, 149-159.

Velicer, W. F., & Fava, J. L. (1987). An evaluation of the effects of variable sampling on component, image, and factor analysis. *Multivariate Behavioral Research*, **22**, 193-210.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting and appropriate procedure. *Multivariate Behavioral Research*, **25**, 1-28.

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

Wakeling, I. N., & Macfie, H. J. (1992). A robust PLS procedure. *Journal of Chemometrics*, **6**, 189-198.

Walczak, B., & Massart, D. L. (1995). Robust principal components regression as a detection tool for outliers. *Chemometrics and Intelligent Laboratory Systems*, **27**, 41-54.

Winsberg, S., & Ramsay, J. O. (1980). Monotonic transformations to additivity using splines. *Biometrika*, **67**, 669-674.

Winsberg, S., & Ramsay, J. O. (1983). Monotonic spline transformations for dimension reduction. *Psychometrika*, **48**, 403-423.

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

Wold, H. (1966). Estimation of principal components and related methods by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 391-420). New York: Academic Press.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modeling: Some current developments. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 383-487). New York: Academic Press.

Wold, H. (1975). Path models with latent variables: the NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Cappecchi (Eds.), *Quantitative Sociology: International Perspectives on Mathematical Statistical Model Building* (pp. 307-357). New York: Academic Press.

Wold, H. (1981). *The Fixed Point Approach to Interdependent Systems*. Amsterdam: North Holland.

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, Part* 2 (pp. 1-54). Amsterdam: North-Holland.

Wold, H. (1985). Partial least squares. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 8 (pp. 481-491). New York: Wiley.

Yanai, H. (1986). Some generalizations of correspondence analysis in terms of projection operators. In E. Diday, Y. Escoufier, L. Lebart, J. P. Pagès, Y. Schektman, & R. Thomassone (Eds.), *Data Analysis and Informatics IV.* (pp. 193-207).

Amsterdam: North Holland.

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

Young, F. W., de Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 505-528.