Structural Equation Modeling: Present and Future

A Festschrift in honor of Karl Jöreskog

Robert Cudeck
Stephen du Toit
Dag Sörbom

Editors

SSI SCIENTIFIC SOFTWARE INTERNATIONAL

FINAL COPY

Structural Equation Modeling: Present and Future. A Festschrift in honor of Karl Jöreskog.

SSI Production Editor: Leo Stam

Cover based on an architectural detail from Frank Lloyd Wright's Coonley House.

1 2 3 4 5 6 7 8 9 0 03 02 01 00

Published by:

Scientific Software International, Inc.
7383 North Lincoln Avenue, Suite 100
Lincolnwood, IL 60712–1704
Tel: +1.847.675.0720
Fax: +1.847.675.2140
URL: http://www.ssicentral.com

ISBN: 0–89498–049–1 (paper)
ISBN: 0–89498–050–5 (cloth)

Contents

FINAL COPY

1 Structural equation modeling for experimental data

Yutaka Kano[1]

Abstract

We first review the use of Structural Equation Modeling (SEM) for the analysis of experimental data. Typical examples include ANOVA, ANCOVA, and MANOVA with or without a covariance structure. SEM for those experimental data is a mean and covariance structure model in multiple populations with a common covariance matrix. Such analyses can be implemented under the assumption that all observed variables be distributed as normal including fixed-effect exogenous variables, which denote levels of factors, for example. Theoretical basis for the usage, based on conditional (likelihood) inference, is explicitly explained.

A bias of a path coefficient estimate particularly in standardized solutions is pointed out which comes from the fact that variance estimates of dependent variables contain variation of means.

Statistical power of several testing procedures concerning mean vectors across several populations are examined, when a factor model can be assumed for observed variables. The procedures considered here are MANOVA, a mean and covariance structure model implemented by SEM, and ANOVA of a factor score or a weighted sum of observed variables. SEM is shown to be the most powerful tool in this context.

[1] This paper is dedicated to the 65th birthday of Prof. Karl G. Jöreskog.

1.1 Introduction

Structural Equation Modeling (SEM) is a very powerful tool for analysis of correlational or observational data, and it can be used for experimental data as well. A typical example is the analysis of multiple populations, originated by Jöreskog (1971b). SEM can be applied to more complex or more typical experimental designs such as ANOVA, ANCOVA, and MANOVA (*e.g.*, Bagozzi 1977; Bagozzi & Yi 1989; Kühnel 1988). Sörbom (1978)'s work on ANCOVA is significant. See also a LISREL manual (*e.g.*, Jöreskog & Sörbom, 1989, pp. 112–116; 1996a, pp. 151–155). To implement these analyses for experimental data one can take a multiple population approach and/or a regression approach. In the regression approach, a design matrix is given explicitly in a data file and draw paths from control variables to dependent variables. It is known that one could treat independent (exogenous) binary variables as if they were continuous under certain assumptions in SEM. Background for these analyses is based on *conditional* statistical inference.

In Section 1.2, we provide some examples of models for experimental data from traditional to modern models that can be analyzed only with the SEM approach. One-factor MANOVA design is employed as a comprehensive example to present the idea of conditional inference. A full explanation for conditional inference will be given in Section 1.3. Equivalence between normal theory inference with exogenous fixed-effect variables and analysis of multiple populations with different mean structures and a common covariance structure is presented.

Section 1.4 gives a cautionary note on the bias of a standardized estimate. In Section 1.5, we compare the power of several statistical tests concerning mean vectors theoretically, and the SEM approach is shown to be most favorable.

1.2 Examples

We shall begin with a simple ANOVA or MANOVA example. Consider six observed dependent variables $Y_1, \ldots, Y_6$ and a binary variable X. Let us say that the Y variables are psychological tests and that X denotes sex with values 0 for male and 1 for female. One would like to examine an effect of sex on the six variables. While the aim is achieved by the traditional MANOVA design, with SEM one can also make the inference by drawing the path diagram as in Fig. 1.1.

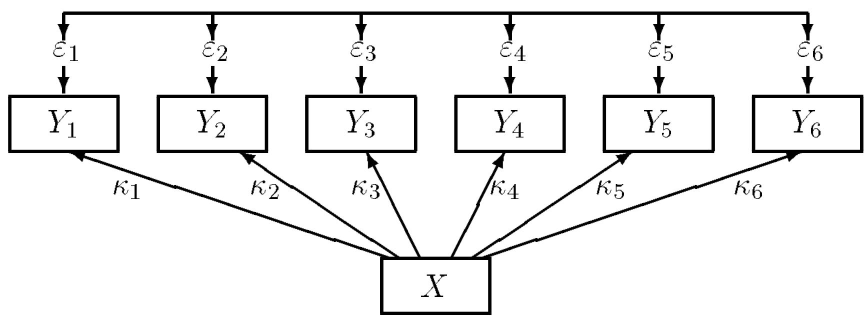

Figure 1.1 ANOVA and MANOVA

The model in Fig. 1.1 is representable in model equations and variance-covariances of independent variables in the form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{bmatrix} X + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}, \quad \left\{ \begin{array}{rcl} E(X) & = & \tau \\ V(X) & = & \phi \\ Cov(\epsilon_i, \epsilon_j) & = & \theta_{ij} \\ & (i, j = 1, \ldots, 6) & \end{array} \right. \tag{1.1}$$

The mean of error variables ϵ_i is assumed implicitly to be zero here and in the sequel as well.

Although variable means are specified in (1.1) it does not imply use of a mean and covariance structure model, because the mean vector has the saturated structure, so that every population mean is estimated by its sample counterpart.

The chi-square difference test between the model in Fig. 1.1 and the model with all the κ_i's zero gives an approximate MANOVA test statistic, which are correct at least asymptotically, to detect difference in mean vectors of $[Y_1, \ldots, Y_6]'$ between males and females.

One usually introduces $q - 1$ dummy variables if X is a treatment variable with q levels. See Jöreskog & Sörbom (1996a, pp. 151–155).

The conditional approach is a useful way to validate the use of independent binary or non-normal variates in many statistical models including SEM.[2] In the approach, the conditional density function given X is considered as the likelihood function to be maximized. We shall denote the equation in (1.1) in matrix notation as $\mathbf{Y} = \boldsymbol{\nu} + \boldsymbol{\kappa}X + \mathbf{e}$ and $\Theta = (\theta_{ij})$.

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_0}$ and $\mathbf{Y}_{n_0+1}, \ldots, \mathbf{Y}_{n_0+n_1}$ be a random sample from the populations of males and females, respectively, and let X_i be a sample that denotes the sex of the i-th observation. Using (1.1), the conditional likelihood (CL) given X_i is

$$\prod_{i=1}^{n_0+n_1} N(\mathbf{Y}_i|\boldsymbol{\nu} + \boldsymbol{\kappa}X_i, \Theta) =$$
$$\prod_{i=1}^{n_0} N(\mathbf{Y}_i|\boldsymbol{\nu} + \boldsymbol{\kappa}x^{(0)}, \Theta) \times \prod_{i=n_0+1}^{n_0+n_1} N(\mathbf{Y}_i|\boldsymbol{\nu} + \boldsymbol{\kappa}x^{(1)}, \Theta) \quad (1.2)$$

with $x^{(0)} = 0$ and $x^{(1)} = 1$. The first term in (1.2) is the likelihood based on the male data whereas the second one is that based on the female data. The difference of the mean vectors is $(\boldsymbol{\nu}+\boldsymbol{\kappa}x^{(1)}) - (\boldsymbol{\nu}+\boldsymbol{\kappa}x^{(0)}) = \boldsymbol{\kappa}$, so that the regression coefficients $\boldsymbol{\kappa}$ represent the difference in the mean vectors between males and females. The error covariance matrix Θ denotes the covariances between the observed variables $\mathbf{Y}$. Under the null hypothesis $H_0 : \boldsymbol{\kappa} = \mathbf{0}$, both populations have the common mean vector $\boldsymbol{\nu}$.

As a result, the inference based on the conditional likelihood in (1.2) is nothing but the usual MANOVA to find the difference between mean vectors. It should be noted that the covariance matrices of the male population and the female population are identical with each other. The ANOVA for testing the difference in each $E(Y_i)$ can be performed by the Wald test as $\hat{\kappa}_i/\widehat{\mathrm{SE}}$, where the SE is the asymptotic standard error of $\hat{\kappa}_i$.

We shall clarify how the conditional approach is related to the full likelihood (FL) approach in which the distribution of $[\mathbf{Y}', X]'$ is regarded erroneously as multivariate normal. The likelihood assuming full normality

[2] The conditional approach and conditional inference used here are different from those notions developed by R.A. Fisher in which the conditional likelihood inference given an ancillary statistic is made.

for $(\mathbf{Y}_i, X_i)$ is

$$\prod_{i=1}^{n_0+n_1} N(\mathbf{Y}_i|\boldsymbol{\nu}+\boldsymbol{\kappa}X_i,\Theta)N(X_i|\tau,\phi)\,. \tag{1.3}$$

The maximum likelihood estimates for $\boldsymbol{\nu}$, $\boldsymbol{\kappa}$, $\boldsymbol{\Theta}$ based on (1.2) coincide with those based on (1.3) because (τ,ϕ) is functionally independent of $\boldsymbol{\nu}$, $\boldsymbol{\kappa}$, $\boldsymbol{\Theta}$. The likelihood (1.3) can be rewritten as

$$\prod_{i=1}^{n_0+n_1} N\left(\begin{bmatrix}\mathbf{Y}_i\\ X_i\end{bmatrix}\middle|\begin{bmatrix}\boldsymbol{\nu}+\boldsymbol{\kappa}\tau\\ \tau\end{bmatrix},\begin{bmatrix}\boldsymbol{\kappa}\phi\boldsymbol{\kappa}'+\Theta & \boldsymbol{\kappa}\phi\\ \phi\boldsymbol{\kappa}' & \phi\end{bmatrix}\right)\,.$$

The population mean $[\boldsymbol{\nu}'+\tau\boldsymbol{\kappa}',\tau]'$ is estimated with the sample averages $[\bar{\mathbf{Y}}',\bar{X}]$ for any value of $\boldsymbol{\kappa}$. One can estimate the covariance matrix $\begin{bmatrix}\boldsymbol{\kappa}\phi\boldsymbol{\kappa}'+\Theta & \boldsymbol{\kappa}\phi\\ \phi\boldsymbol{\kappa}' & \phi\end{bmatrix}$, which is based on the path diagram in Fig. 1.1, by the usual sample covariance matrix of $[\mathbf{Y}', X_i]$.

It will be deduced from the discussion above that the normal theory analysis based on the usual covariance structure model, without mean structures, defined by the path diagram in Fig. 1.1 is equivalent to the analysis of multiple populations with different mean structures and a common covariance structure.

A model for MANOVA with two treatments A and B with interaction is expressible in path diagram form as shown in Fig. 1.2.

The independent variables X_1, X_2, X_3 denote effects of the treatments of A, B and their interaction $A\times B$ and take values for every combination of levels of A and B as

	X_1	X_2	X_3
A_1B_1	1	1	1
A_2B_1	−1	1	−1
A_1B_2	1	−1	−1
A_2B_2	−1	−1	1

The model in Fig. 1.2 can be represented as

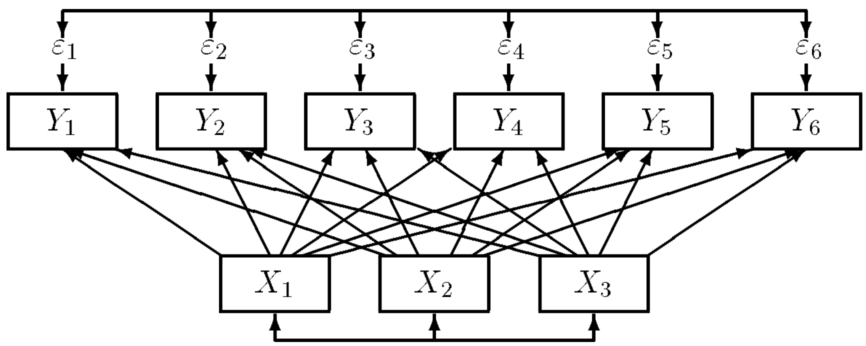

Figure 1.2 ANOVA and MANOVA with two treatments

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \\ \kappa_{41} & \kappa_{42} & \kappa_{43} \\ \kappa_{51} & \kappa_{52} & \kappa_{53} \\ \kappa_{61} & \kappa_{62} & \kappa_{63} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}, \tag{1.4}$$

$$\begin{cases} E(X_i) & = & \tau_i \ (i = 1, 2, 3) \\ Cov(X_r, X_s) & = & \phi_{rs} \ (r, s = 1, 2, 3) \\ Cov(\epsilon_i, \epsilon_j) & = & \theta_{ij} \ (i, j = 1, \ldots, 6) \end{cases} .$$

There is a situation where a certain model can be assumed to explain covariances between dependent variables $\mathbf{Y}$. The model in Fig. 1.3 assumes a two-factor model, and the mean vectors can differ by $\boldsymbol{\kappa}$ (path coefficients from X_1 to Y_i) between males and females.

The equations and variance-covariances of the model in Fig. 1.3 are as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{bmatrix} X_1 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}, \tag{1.5}$$

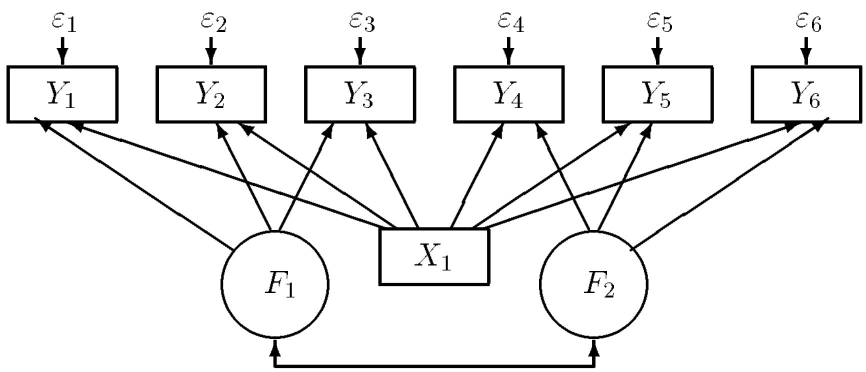

Figure 1.3
MANOVA with a factor analysis model for covariance matrices

$$\left\{ \begin{array}{rcl} E(X_1) & = & \tau_1, \quad V(X_1) = \phi_1 \\ Cov(F_r, F_s) & = & \psi_{rs} \quad (r, s = 1, 2) \\ V(\epsilon_i) & = & \theta_{ii} \quad (i = 1, \ldots, 6) \end{array} \right. .$$

In the model in Fig. 1.3, $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.

It is occasionally realistic to assume there is difference in the mean of latent variables between males and females, which results in a mean difference of observed variables $\mathbf{Y}$. The model is expressible as in Fig. 1.4. The equations and variance-covariances of the model in Fig. 1.4 are as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} X_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}, \tag{1.6}$$

$$\left\{ \begin{array}{rcl} E(X_1) & = & \tau_1, \ V(X_1) = \phi_1 \\ Cov(\zeta_i, \zeta_j) & = & \psi_{ij} \ (i, j = 1, 2) \\ V(\epsilon_i) & = & \theta_{ii} \ (i = 1, \ldots, 6) \end{array} \right. .$$

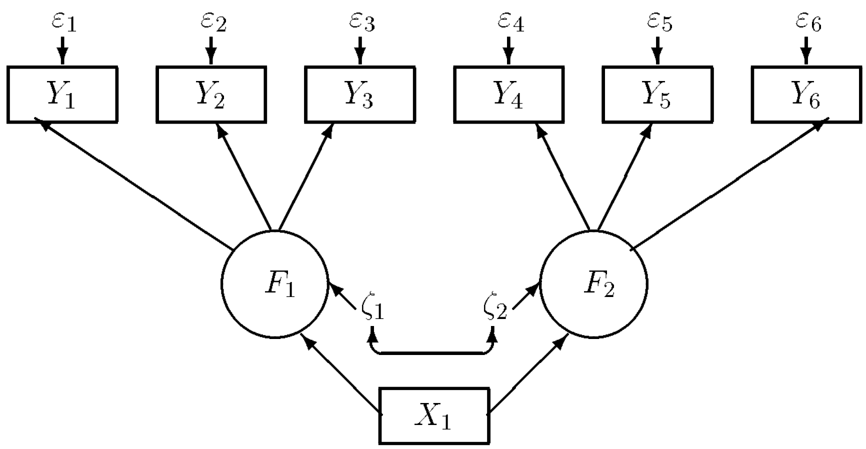

Figure 1.4 A model for mean difference between male and female

The model is a kind of MIMIC model. Jöreskog & Goldberger (1975) studied statistical properties of the MIMIC model with fixed regressors (conditioned on $\mathbf{X}$) and with random regressors, and obtained the same fitting functions for both MIMIC model specifications. Muthén (1989a) emphasized the use of MIMIC models to express heterogeneity in means in several groups.

We shall end this section by giving an example of ANCOVA for a latent variable. Sörbom (1978) developed an alternative model to the MANCOVA in which a latent variable was introduced and the modeling allowed for heterogeneous covariance matrices. Sörbom's model uses multi-sample analysis of mean and covariance structures. Arbuckle & Wothke (1999, Example 9) show one sample covariance structure modeling in the case where the covariance matrices are homogeneous. This modeling is presented in Fig. 1.5. Let us say that Y_1 to Y_3 and Y_4 to Y_6 are pre-tests and post-tests on verbal ability, and X_1 is a binary variable that denotes whether the group is control or experimental.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

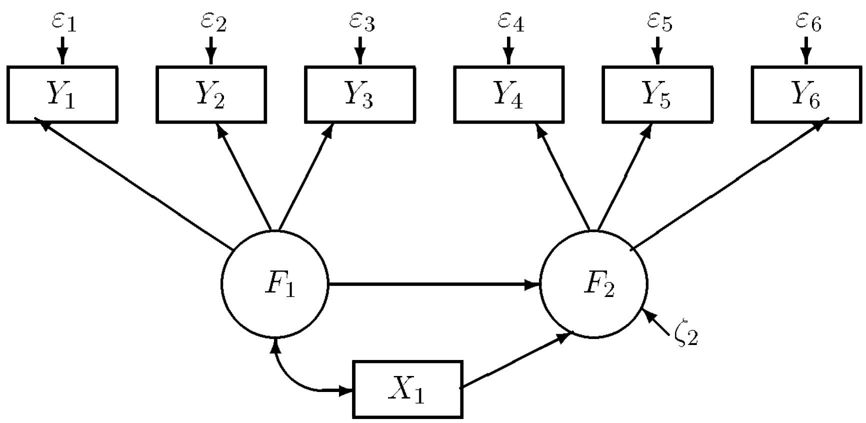

Figure 1.5 **ANCOVA for a latent variable**

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \boldsymbol{\beta}_{21} & 0 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma_2 \end{bmatrix} X_1 + \begin{bmatrix} F_1 \\ \zeta_2 \end{bmatrix}, \tag{1.7}$$

$$\left\{ \begin{array}{rcl} E(X_1) & = & \tau_1, \ \ V(X_1) = \Phi \\ V\left(\begin{bmatrix} X_1 \\ F_1 \\ \zeta_2 \end{bmatrix}\right) & = & \begin{bmatrix} \phi_{11} & \phi_{12} & 0 \\ \phi_{21} & \phi_{22} & 0 \\ 0 & 0 & \psi_{22} \end{bmatrix} \\ V(\epsilon_i) & = & \theta_{ii} \ \ (i = 1, \ldots, 6) \end{array} \right.$$

Note that the first row of the second equation, $F_1 = F_1$, is redundant, but it will be useful to construct a covariance structure of observed variables (it is a kind of RAM representation; see McArdle & McDonald 1984).

1.3 Conditional inference

All the models in Section 1.2 can be expressed as

$$\begin{array}{rcl} \mathbf{Y} & = & \boldsymbol{\nu} + \Lambda \mathbf{f} + K\mathbf{X} + \boldsymbol{\epsilon} \\ \mathbf{f} & = & \boldsymbol{\alpha} + B\mathbf{f} + \Gamma \mathbf{X} + \boldsymbol{\zeta} \end{array}$$

$$\left\{ \begin{array}{rcl} E(\mathbf{X}) & = & \boldsymbol{\tau} \\ E(\boldsymbol{\epsilon}) & = & \mathbf{0} \\ E(\boldsymbol{\zeta}) & = & \mathbf{0} \end{array} \right., \quad \left\{ \begin{array}{rcl} V(\mathbf{X}) & = & \Phi \\ V(\boldsymbol{\epsilon}) & = & \Theta \\ V(\boldsymbol{\zeta}) & = & \Psi \end{array} \right., \quad \left\{ \begin{array}{rcl} Cov(\mathbf{X}, \boldsymbol{\epsilon}) & = & O \\ Cov(\boldsymbol{\epsilon}, \boldsymbol{\zeta}) & = & O \\ Cov(\boldsymbol{\zeta}, \mathbf{X}) & = & \Sigma_{\zeta X} \end{array} \right. .$$

Here $\mathbf{X}$ and $\mathbf{Y}$ denote independent and dependent observed variables whereas $\mathbf{f}$ is a vector of latent variables (constructs). The $K\mathbf{X}$ and $\Gamma\mathbf{X}$ denote direct effects of $\mathbf{X}$ on $\mathbf{Y}$ and indirect effects of $\mathbf{X}$ through $\mathbf{f}$ on $\mathbf{Y}$, respectively. The diagonals of B are fixed to be zero and $I - B$ is assumed to be nonsingular.

It holds that $\Sigma_{\zeta X} = O$ in all the examples in Section 1.2 with the exception of the model shown in Fig. 1.5. So we first assume that $\Sigma_{\zeta X} = O$ for simplicity.

In SEM, it is convention that a joint normal distribution is assumed for $\mathbf{X}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$, so that the observed vector $[\mathbf{Y}', \mathbf{X}']'$ follows a multivariate normal distribution. Eliminating $\mathbf{f}$ in the equations (1.8), we have

$$\mathbf{Y} = \boldsymbol{\nu} + \Lambda(I - B)^{-1}(\boldsymbol{\alpha} + \boldsymbol{\zeta}) + \boldsymbol{\epsilon} + (\Lambda(I - B)^{-1}\Gamma + K)\mathbf{X}.$$

Under the normal assumption, the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is

$$\mathbf{Y}|\mathbf{X} = \boldsymbol{x} \sim N(\boldsymbol{y}|E(\mathbf{Y}|\boldsymbol{x}), V(\mathbf{Y}|\boldsymbol{x})),$$

with

$$\begin{aligned} E(\mathbf{Y}|\boldsymbol{x}) &= \boldsymbol{\nu} + \Lambda(I - B)^{-1}\boldsymbol{\alpha} + (\Lambda(I - B)^{-1}\Gamma + K)\boldsymbol{x} \\ V(\mathbf{Y}|\boldsymbol{x}) &= \Lambda(I - B)^{-1}\Psi(I - B')^{-1}\Lambda' + \Theta \ \ (= \Sigma_{Y \cdot X}, \text{ say}). \end{aligned} \qquad (1.8)$$

Note that $\Sigma_{Y \cdot X}$ is independent of $\boldsymbol{x}$. If $\Sigma_{\zeta X} \neq O$, $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ have to be replaced, respectively, with

$$\begin{aligned} \boldsymbol{\alpha} + E(\boldsymbol{\zeta}|\boldsymbol{x}) &= \boldsymbol{\alpha} + \Sigma_{\zeta X}\Phi^{-1}(\boldsymbol{x} - \boldsymbol{\tau}) \qquad \text{and} \\ V(\boldsymbol{\zeta}|\boldsymbol{x}) &= \Psi - \Sigma_{\zeta X}\Phi^{-1}\Sigma_{X\zeta} \, . \end{aligned} \qquad (1.9)$$

The normal theory inference assumes

$$[\mathbf{Y}', \mathbf{X}']' \sim N(\boldsymbol{y}|E(\mathbf{Y}|\boldsymbol{x}), \Sigma_{Y \cdot X})N(\boldsymbol{x}|\boldsymbol{\tau}, \Phi),$$

and the normal theory MLE is a solution that maximizes the likelihood using the density above. We will call it the conditional inference that the only conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is specified and no particular distributional assumption is made on $\mathbf{X}$, that is,

$$\mathbf{Y}|\mathbf{X} = \boldsymbol{x} \sim N(\boldsymbol{y}|E(\mathbf{Y}|\boldsymbol{x}), \Sigma_{Y \cdot X}).$$

There is a close relationship between the normal theory inference and the conditional inference, as explained below.

Instead of a normality assumption on the entire observation, the conditional inference assumes that (i) the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is multivariate normal with $E(\mathbf{Y}|\mathbf{X})$ and $V(\mathbf{Y}|\mathbf{X})$ in (1.8), (ii) Θ and $\mathbf{\Psi}$ are unrelated to the value of $\mathbf{X}$ and (iii) the parameters $\boldsymbol{\tau}$ and Φ of $\mathbf{X}$ are not restricted and functionally unrelated with the other parameters. The assumptions (i) and (ii) are on the distribution of the observations; whereas (iii) is related to parameterization of the model considered.

The purpose of this section is to study what happens to the normal theory statistical inference, if $\mathbf{X}$ is not normally distributed.

It is obvious that the assumptions (i) and (ii) hold for the normal distribution for all the variables. Consider non-normal distributions for $\mathbf{X}$. If $\mathbf{X}$ is distributed independently of $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$ or $\mathbf{X}$ is a fixed variable, (ii) is met. This model has been developed by Muthén (1984, 1989a) and Muthén & Muthén (1998). It is an excellent idea, due to Muthén, to split observed variables into independent and dependent variables.

We shall denote by $\boldsymbol{\pi}$ all the parameters involved in $E(\mathbf{Y}|\mathbf{X})$ and $V(\mathbf{Y}|\mathbf{X})$. By Assumption (iii), $\boldsymbol{\pi}$ is functionally independent of $(\boldsymbol{\tau}, \Phi)$.

Suppose that the distribution of $\mathbf{X}$ is specified as $N(\boldsymbol{x}|\boldsymbol{\tau}, \Phi)$, which may be misspecified. The likelihood on the full data $(\mathbf{Y}_i, \mathbf{X}_i)$ and the likelihood on the conditional data $\mathbf{Y}_i|\mathbf{X}_i$ $(i = 1, 2, \ldots, n)$ are expressed respectively as

$$\mathrm{FL}_N(\boldsymbol{\pi}, \boldsymbol{\tau}, \Phi) = \prod_{i=1}^{n} N\Big(\mathbf{Y}_i \Big| E(\mathbf{Y}|\mathbf{X}_i), \Sigma_{Y\cdot X}\Big) N(\mathbf{X}_i|\boldsymbol{\tau}, \Phi), \quad (1.10)$$

$$\mathrm{CL}(\boldsymbol{\pi}) = \prod_{i=1}^{n} N\Big(\mathbf{Y}_i \Big| E(\mathbf{Y}|\mathbf{X}_i), \Sigma_{Y\cdot X}\Big) \quad (1.11)$$

It follows from the functional independence between $\boldsymbol{\pi}$ and $(\boldsymbol{\tau}, \Phi)$ that

$$\max_{(\boldsymbol{\pi},\boldsymbol{\tau},\mathbf{\Phi})} \mathrm{FL}_N(\boldsymbol{\pi}, \boldsymbol{\tau}, \mathbf{\Phi}) =$$

$$\max_{\boldsymbol{\pi}} \prod_{i=1}^{N} N\Big(\mathbf{Y}_i \Big| E(\mathbf{Y}|\mathbf{X}_i), \Sigma_{Y\cdot X}\Big) \times \max_{(\boldsymbol{\tau},\Phi)} \prod_{i=1}^{N} N(\mathbf{X}_i|\boldsymbol{\tau}, \Phi) =$$

$$\max_{\boldsymbol{\pi}} \mathrm{CL}(\boldsymbol{\pi}) \times \max_{(\boldsymbol{\tau},\Phi)} \prod_{i=1}^{N} N(\mathbf{X}_i|\boldsymbol{\tau}, \Phi)\,. \quad (1.12)$$

This implies that the maximization with respect to $\boldsymbol{\pi}$ can be made independently of the maximization with respect to $(\boldsymbol{\tau}, \Phi)$. The MLE for $\boldsymbol{\pi}$ based on the full normal theory likelihood (1.10) is identical with that based on the conditional likelihood (1.11).

Let the true density or probability function of $\mathbf{X}$ be written as $f_X(\boldsymbol{x}|\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is a parameter vector associated with the distribution $\mathbf{X}$, which may be $(\boldsymbol{\tau}, \Phi)$. When $\mathbf{X}$ is a fixed variable, $f_X(\boldsymbol{x}|\boldsymbol{\omega}) = 1$. The likelihood based on the true and full density is then expressible as

$$\mathsf{FL}(\boldsymbol{\pi}, \boldsymbol{\omega}) = \prod_{i=1}^{n} N\left(\mathbf{Y}_i \middle| E(\mathbf{Y}|\mathbf{X}_i), \Sigma_{Y \cdot X}\right) f_X(\mathbf{X}_i|\boldsymbol{\omega}) \,. \tag{1.13}$$

In the same manner, the independence between $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$ shows

$$\max_{(\boldsymbol{\pi},\boldsymbol{\omega})} \mathsf{FL}(\boldsymbol{\pi}, \boldsymbol{\omega}) = \max_{\boldsymbol{\pi}} \mathsf{CL}(\boldsymbol{\pi}) \times \max_{\boldsymbol{\omega}} \prod_{i=1}^{n} f_X(\mathbf{X}_i|\boldsymbol{\omega}) \,. \tag{1.14}$$

From (1.12) and (1.14) we have that the MLE based on the normal likelihood, conditional likelihood, and the (true) full likelihood coincide with one another. The MLE is free from the distribution of $\mathbf{X}_i$. This is one of the major reasons for the conditional inference to be allowed.

Although the MLE's are identical, their distributions can be different. The distributions do depend on $f_X(\boldsymbol{x}|\boldsymbol{\omega})$. Rather than the distributions of the MLE's themselves, researchers are more interested in the distribution of a statistic to test $H_0 : \pi_i = 0$ with a Wald type statistic or in the confidence interval of π_i, where π_i is any parameter in $\boldsymbol{\pi}$. We know that under the null hypothesis

$$W = \hat{\pi}_i / \sqrt{\widehat{AV}(\hat{\pi}|\mathbf{X}_i\text{'s})} \xrightarrow{L} N(0,1) \,,$$

as $n \to \infty$, where $AV(\hat{\pi}|\mathbf{X}\text{'s})$ represents the conditional asymptotic variance of $\hat{\pi}$ given $\mathbf{X}_i$'s and $\xrightarrow{L}$ denotes convergence in distribution. Since the asymptotic distribution is $N(0,1)$, independent of the conditioning variable $\mathbf{X}_i$, the convergence to the normal distribution will be true for the unconditional distribution. It is unclear whether the Wald statistic remains valid for non-normal or fixed-effect independent variables $\mathbf{X}$ if the asymptotic variance of $\hat{\pi}$ is constructed not under the conditional distribution but under the fully normal assumption. For this case, the following derivation will be more easily understood.

The normal theory likelihood ratio (LR) test, or chi-square difference test, for $H_0 : \pi_i = 0$ is valid for any distribution for $\mathbf{X_i}$ because

$$\frac{\max_{(\boldsymbol{\pi},\boldsymbol{\omega}):\pi_i=0} \mathsf{FL}_N(\boldsymbol{\pi},\boldsymbol{\omega})}{\max_{(\boldsymbol{\pi},\boldsymbol{\omega})} \mathsf{FL}_N(\boldsymbol{\pi},\boldsymbol{\omega})} = \frac{\max_{\boldsymbol{\pi}:\pi_i=0} \mathsf{CL}(\boldsymbol{\pi}) \max_{(\boldsymbol{\tau},\boldsymbol{\Phi})} N(\mathbf{X}_i|\boldsymbol{\tau},\Phi)}{\max_{\boldsymbol{\pi}} \mathsf{CL}(\boldsymbol{\pi}) \max_{(\boldsymbol{\tau},\boldsymbol{\Phi})} N(\mathbf{X}_i|\boldsymbol{\tau},\Phi)}$$

$$= \frac{\max_{\boldsymbol{\pi}:\pi_i=0} \mathsf{CL}(\boldsymbol{\pi}) \max_{\boldsymbol{\omega}} f_X(\mathbf{X}_i|\boldsymbol{\omega})}{\max_{\boldsymbol{\pi}} \mathsf{CL}(\boldsymbol{\pi}) \max_{\boldsymbol{\omega}} f_X(\mathbf{X}_i|\boldsymbol{\omega})} = \frac{\max_{(\boldsymbol{\pi},\boldsymbol{\omega}):\pi_i=0} \mathsf{FL}(\boldsymbol{\pi},\boldsymbol{\omega})}{\max_{(\boldsymbol{\pi},\boldsymbol{\omega})} \mathsf{FL}(\boldsymbol{\pi},\boldsymbol{\omega})} .$$

The validity of the normal theory Wald test is deduced from the validity of the normal theory LR test and a general theorem of equivalence between the Wald test and LR test (see, *e.g.*, Buse 1982).

Finally, we shall consider the chi-square LR statistic for testing goodness of fit of the model considered. There is difficulty in making such a clear description for the behavior of the chi-square statistic as that for the MLE and the Wald test in the conditional inference. It is very difficult to see how to specify the saturated model for non-normal cases. Here we take as a saturated model

$$N(\boldsymbol{y}|\boldsymbol{\nu}+K\boldsymbol{x},\Sigma)f_X(\boldsymbol{x}|\boldsymbol{\omega}),$$

where $\boldsymbol{\nu}$, K, Σ and $\boldsymbol{\omega}$ consist of all free parameters. Note that $N(\boldsymbol{y}|\boldsymbol{\nu}+K\boldsymbol{x},\Sigma)N(\boldsymbol{x}|\boldsymbol{\tau},\Phi)$ is the saturated model for $(\mathbf{Y},\mathbf{X})$. One important assumption here is that the modeling for $\mathbf{X}$ is the same for the null and saturated models. The normal theory LR statistic for testing goodness of fit is then equivalent to the LR statistic based on the true model specifying $f_X(\boldsymbol{x}|\boldsymbol{\omega})$ for the $\mathbf{X}_i$ because

$$\frac{\max_{(\boldsymbol{\pi},\boldsymbol{\tau},\Phi)} \prod_{i=1}^n N\left(\mathbf{Y}_i \middle| E(\mathbf{Y}_i|\mathbf{X}_i),\Sigma_{Y\cdot X}\right) N(\mathbf{X}_i|\boldsymbol{\tau},\Phi)}{\max_{(\boldsymbol{\nu},K,\Sigma,\boldsymbol{\tau},\Phi)} \prod_{i=1}^n N\left(\mathbf{Y}_i \middle| \boldsymbol{\nu}+K\mathbf{X}_i,\Sigma\right) N(\mathbf{X}_i|\boldsymbol{\tau},\Phi)} =$$

$$\frac{\max_{\boldsymbol{\pi}} \prod_{i=1}^n N\left(\mathbf{Y}_i \middle| E(\mathbf{Y}_i|\mathbf{X}_i),\Sigma_{Y\cdot X}\right)}{\max_{(\boldsymbol{\nu},K,\Sigma)} \prod_{i=1}^n N\left(\mathbf{Y}_i \middle| \boldsymbol{\nu}+K\mathbf{X}_i,\Sigma\right)} =$$

$$\frac{\max\limits_{(\boldsymbol{\pi},\boldsymbol{\omega})} \prod_{i=1}^{n} N\left(\mathbf{Y}_i \middle| E(\mathbf{Y}_i|\mathbf{X}_i), \Sigma_{Y\cdot X}\right) f_X(\mathbf{X}_i|\boldsymbol{\omega})}{\max\limits_{(\boldsymbol{\nu},K,\Sigma,\boldsymbol{\omega})} \prod_{i=1}^{n} N\left(\mathbf{Y}_i \middle| \boldsymbol{\nu} + K\mathbf{X}_i, \Sigma\right) f_X(\mathbf{X}_i|\boldsymbol{\omega})}. \tag{1.15}$$

Again, the key assumptions for the normal model to be relevant are that $\boldsymbol{\tau}$ and Φ are unrestricted and are functionally independent of the parameters $\boldsymbol{\pi}$. As a result, the normal theory statistical inference such as point estimation, Wald test and goodness of fit test are all correct irrespective of any type of distribution of independent observed variables $\mathbf{X}$, provided that $E(\mathbf{X})$ and $V(\mathbf{X})$ are all free parameters and they are functionally independent of the other parameters $\boldsymbol{\pi}$.

From now on, we shall consider the case where $\Sigma_{\zeta X}$ may not be zero. In this case, $E(\mathbf{Y}|\mathbf{X})$ and $V(\mathbf{Y}|\mathbf{X})$ involves $\boldsymbol{\tau}$ and Φ, so that $\boldsymbol{\pi}$ is functionally dependent on $(\boldsymbol{\tau}, \Phi)$. However, the derivation above is also applicable if there is a suitable parameter transformation such that the parameters in $E(\mathbf{Y}|\mathbf{X})$ and $V(\mathbf{Y}|\mathbf{X})$ are functionally independent of $(\boldsymbol{\tau}, \Phi)$. Whether there is such a transformation will depend on the model under consideration. In the model in Fig. 1.5 with (1.7), we have

$$\begin{aligned} E(\boldsymbol{\zeta}|X_1) &= E\left(\begin{bmatrix} F_1 \\ \zeta_2 \end{bmatrix} \middle| X_1\right) = \begin{bmatrix} \phi_{21}\phi_{11}^{-1}(X_1 - \tau_1) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma_2 X_1 \end{bmatrix} \\ V(\boldsymbol{\zeta}|X_1) &= V\left(\begin{bmatrix} F_1 \\ \zeta_2 \end{bmatrix} \middle| X_1\right) = \begin{bmatrix} \phi_{22} - \phi_{21}\phi_{11}^{-1}\phi_{12} & 0 \\ 0 & \psi_{22} \end{bmatrix}. \end{aligned}$$

It is easily understood that $V(\boldsymbol{\zeta}|X_1)$ can be written as $\begin{bmatrix} \phi_{22}^* & 0 \\ 0 & \psi_{22} \end{bmatrix}$ with ϕ_{22}^* a free parameter functionally independent of (τ_1, ϕ_{11}), because ϕ_{22} is a free parameter in $\boldsymbol{\pi}$. Since X_1 takes a value of 0 or 1, the conditional expectation $E(\mathbf{Y}|X_1)$ can be expressed as

$$\begin{aligned} E(\mathbf{Y}|X_1 = 0) &= \boldsymbol{\nu} + \Lambda(I - B)^{-1} \begin{bmatrix} \phi_{21}\phi_{11}^{-1}(0 - \tau_1) \\ 0 \end{bmatrix} \\ E(\mathbf{Y}|X_1 = 1) &= \boldsymbol{\nu} + \Lambda(I - B)^{-1} \begin{bmatrix} \phi_{21}\phi_{11}^{-1}(1 - \tau_1) \\ \gamma_2 \end{bmatrix}. \end{aligned}$$

The terms involving τ_1 can be absorbed into $\boldsymbol{\nu}$ (so that it is written as $\boldsymbol{\nu}^*$) and $\phi_{21}\phi_{11}^{-1}$ can be written as ϕ_{21}^*, a parameter in $\boldsymbol{\pi}$ which is independent of (τ_1, ϕ_{11}). We thus have

$$\begin{aligned} E(\mathbf{Y}|X_1 = 0) &= \boldsymbol{\nu}^* + \Lambda(I - B)^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ E(\mathbf{Y}|X_1 = 1) &= \boldsymbol{\nu}^* + \Lambda(I - B)^{-1} \begin{bmatrix} \phi_{21}^* \\ \gamma_2 \end{bmatrix} . \end{aligned}$$

In summary, the argument above shows that the model in Fig. (1.5) assuming normality for $[\mathbf{Y}', \mathbf{X}']'$ is equivalent to the joint analysis of the two populations for $\mathbf{Y}$, corresponding to the values of X_1, assuming the common covariance structure but different mean structures, that is, $E(F_1) = E(F_2) = 0$ for $X_1 = 0$ and $E(F_1)$ and $E(F_2)$ being free parameters for $X_1 = 1$. The difference in $E(F_1)$ between the two populations indicates failure of random assignment of subjects to the control or experimental group.

1.4 Bias

Consider the model in Fig. 1.4. There are mainly two purposes when applying this kind of model. One is to study impacts of X_1 on the means of the latent variables. The purpose will be achieved by the Wald test for the coefficients γ_i's or the difference chi-square test between this model and the model with γ_i's zero. In this case, the measurement model is not of main interest.

The other purpose is to introduce X_1 to express heterogeneity of means in the sample and to study factor structure of the observed variables after adjusting the mean differences. Such usage was emphasized by Muthén (1989). In this case, the estimates in the measurement model are important. The model is said to be a MIMIC model. In this section, we point out a bias of factor loading estimates in a measurement model and suggest a formula for correction. The problem is important for the second purpose above.

Now we illustrate the bias of an estimate using a simple example. Fig. 1.6 and 1.7 show unstandardized and standardized estimates, respectively, in a MIMIC model. Let us say that X_1 is a binary variable to represent sex and that one is interested in the factor structure after adjusting factor mean difference between males and females. According to Fig. 1.6, the

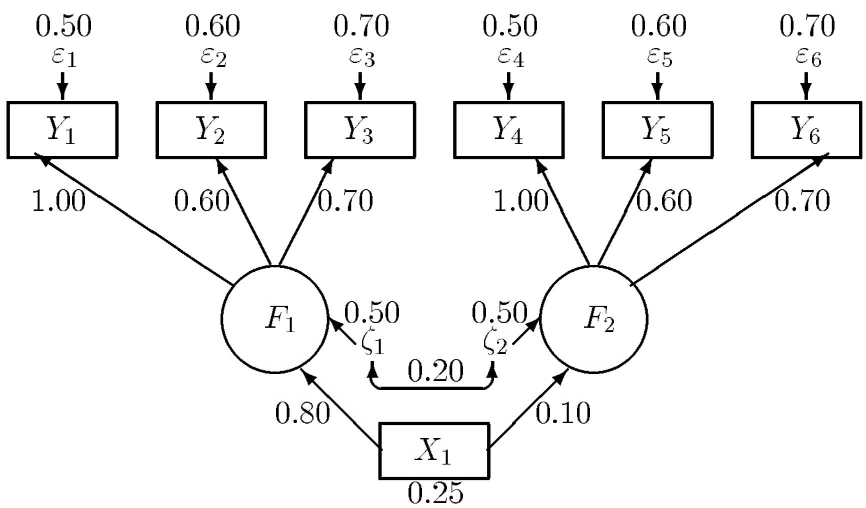

Figure 1.6
Factor analysis with different latent means: Unstandardized estimates

latent mean differences are 0.80 for F_1 and 0.1 for F_2. Since the factor variances, equal to those of ζ_i, are 0.5 for both factors, the standardized estimates in the factor structure should be those in Fig. 1.8. In the example, the measurement model for F_1 is the same as that for F_2. However, usual standardized estimates given in Fig. 1.7 are not equivalent to those in Fig. 1.8. The measurement models for F_1 and F_2 are not identical. A bias certainly arises in the model in Fig. 1.7.[3]

The bias comes from the fact that the factor variances are calculated as if X_1 were treated as a random-effect variable, namely

$$\begin{aligned} V(F_1) &= V(0.80X_1 + \zeta_1) = 0.1600 + 0.50 = 0.6600 \\ V(F_2) &= V(0.10X_1 + \zeta_2) = 0.0025 + 0.50 = 0.5025\ . \end{aligned}$$

The normal theory standardized estimates in Fig. 1.7 have then been calculated as:

$$0.6\sqrt{\frac{V(F_1)}{V(Y_2)}} = 0.6\sqrt{\frac{0.6600}{0.6^2 \times 0.6600 + 0.6}} = 0.5326 \quad \text{for} \quad \lambda_{21}$$

[3]The path coefficients from F_2 are different between Fig. 1.7 and Fig. 1.8 but appear to be the same within rounding error. It happens because the impact of X_1 on F_2 is small.

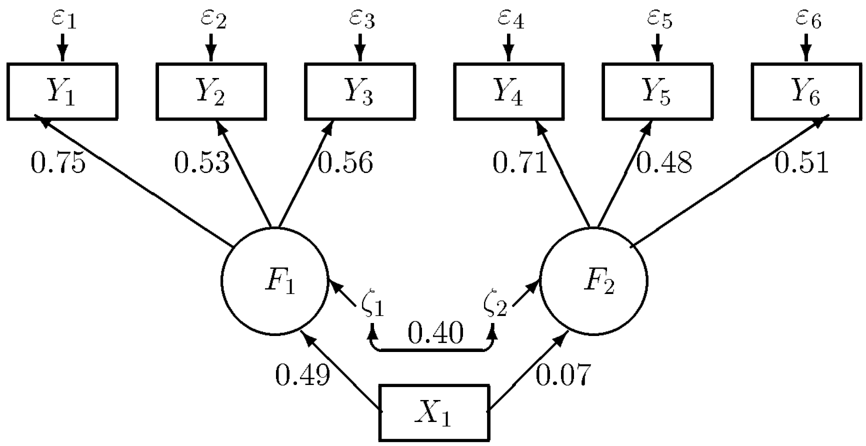

Figure 1.7
Factor analysis with different latent means: Standardized estimates

$$0.6\sqrt{\frac{V(F_1)}{V(Y_5)}} = 0.6\sqrt{\frac{0.5025}{0.6^2 \times 0.5025 + 0.6}} = 0.4813 \quad \text{for} \quad \lambda_{52}\ ,$$

for example. On the contrary, the corresponding true standardized estimates in Fig. 1.8 are

$$0.6\sqrt{\frac{0.50}{0.6^2 \times 0.50 + 0.6}} = 0.4803 \quad \text{for both} \quad \lambda_{21} \quad \text{and} \quad \lambda_{52}\ .$$

Such a bias arises in the model in Fig. 1.3 as well. The data that can be analyzed with the model in Fig. 1.6 can also be analyzed with the model in Fig. 1.3. Estimates unstandardized and standardized in a usual way are given in Table 1.1. Again, the standardized estimates are different from what is expected.

We do not mention that unstandardized estimates should be reported in the case where a mean adjustment is made with independent fixed-effect variables. One should be careful to figure out the variance of a variable that is influenced directly and/or indirectly by independent fixed-effect variables. The correct variance can be calculated by deducting the fixed-effect variable variances from the formal variance obtained by normal theory. Standardized solutions are then calculated using the true variances (rather than the formal one) in the usual manner.

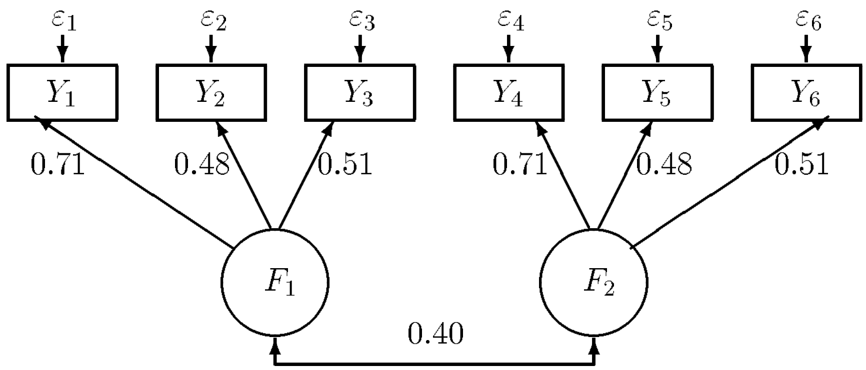

Figure 1.8
Factor analysis after adjusting different latent means: Standardized estimates

Table 1.1 Estimates of path coefficient

	unstandardized			*standardized*		
	F_1	F_2	X_1	F_1	F_2	X_1
Y_1	1.00		0.80	0.66		0.37
Y_2	0.60		0.48	0.46		0.26
Y_3	0.70		0.56	0.49		0.78
Y_4		1.00	0.10		0.71	0.05
Y_5		0.60	0.06		0.48	0.03
Y_6		0.70	0.07		0.51	0.04

1.5 Statistical power

In this section, we shall make comparisons of statistical power among several procedures to test mean differences of manifest variables or latent variables. Consider a MIMIC model in Fig. 1.9. The fixed-effect variable X_1 takes on values of 0 or 1. Without the effect of X_1, the model would be a one-factor model with covariance matrix $\Sigma = \boldsymbol{\lambda}\phi\boldsymbol{\lambda}' + \Psi$, where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]'$, $V(F_1) = \phi$ and $V([\epsilon_1, \ldots, \epsilon_p]') = \Psi$ with $p = 6$. The conditional distribution of $\mathbf{Y} = [Y_1, \ldots, Y_p]'$ given X_1 is $N_p(\boldsymbol{\nu}, \Sigma)$ for $X_1 = 0$ and $N_p(\boldsymbol{\nu} + \boldsymbol{\lambda}\gamma, \Sigma)$ for $X_1 = 1$.

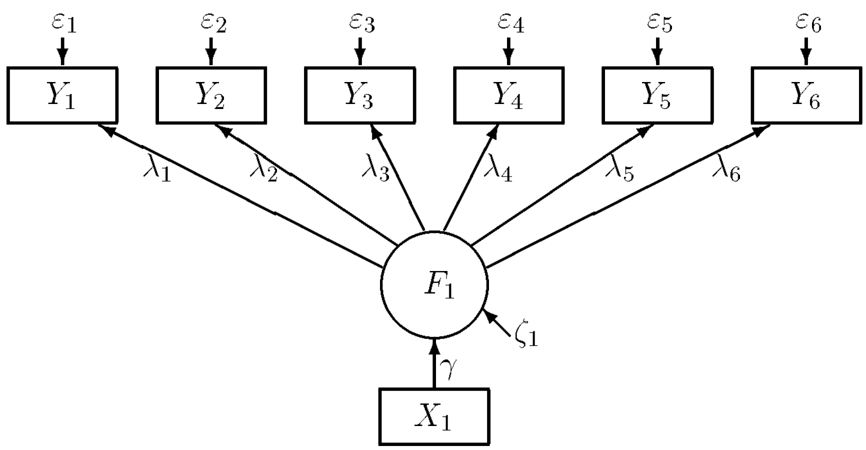

Figure 1.9 MIMIC model

There are many ways to test statistically the effect of X_1. A traditional method is a MANOVA. The MANOVA may not be so powerful because it ignores the mean structure $\boldsymbol{\lambda}\gamma$. In fact, Hancock, Lawrence, & Nevitt (2000) showed empirically that the MANOVA is less powerful than the MIMIC approach. The second procedure considered here is to use the SEM approach to test $\gamma = 0$. An alternative way often used is to sum up the observed variables and to make an ANOVA to compare $E(\sum_{i=1}^{p} Y_i)$ between the populations for $X_1 = 0$ and for $X_1 = 1$. One can use factor scores; that is, to compare $E(\boldsymbol{\lambda}'\Sigma^{-1}\mathbf{Y})$ (or $E[(\boldsymbol{\lambda}'\Psi^{-1}\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\Psi^{-1}\mathbf{Y}]$) between the two populations. The latter two approaches can be considered as a comparison of $E(\mathbf{c}'\mathbf{Y})$ with $\mathbf{c}$ a constant vector.

To simplify mathematics needed to compare those procedures, we assume $\boldsymbol{\lambda}$, ϕ and $\boldsymbol{\Psi}$ (and hence Σ) are known. Then $\boldsymbol{\nu}$ and γ are unknown parameters to be estimated.

Let $\bar{\mathbf{Y}}^{(0)}$ and $\bar{\mathbf{Y}}^{(1)}$ be sample mean vectors with sample size n_0 and n_1, respectively, from the populations with $X_1 = 0$ and $X_1 = 1$.

The MANOVA test statistic is equivalent to

$$\frac{n_0 n_1}{n_0 + n_1}\left(\bar{\mathbf{Y}}^{(0)} - \bar{\mathbf{Y}}^{(1)}\right)' \Sigma^{-1} \left(\bar{\mathbf{Y}}^{(0)} - \bar{\mathbf{Y}}^{(1)}\right) .$$

See *e.g.*, Anderson (1984, Sec. 4). The statistic follows the chi-square distribution of p degrees of freedom and the non-centrality parameter

$$\delta^2_{\text{MANOVA}} = \frac{n_0 n_1 \gamma^2}{n_0 + n_1}\boldsymbol{\lambda}'\Sigma^{-1}\boldsymbol{\lambda} . \tag{1.16}$$

Note that the mean and covariance structure is represented as

$$\begin{aligned} E\begin{bmatrix} \bar{\mathbf{Y}}^{(0)} \\ \bar{\mathbf{Y}}^{(1)} \end{bmatrix} &= \begin{bmatrix} I_p & \mathbf{0} \\ I_p & \boldsymbol{\lambda} \end{bmatrix}\begin{bmatrix} \boldsymbol{\nu} \\ \gamma \end{bmatrix} \\ V\begin{bmatrix} \bar{\mathbf{Y}}^{(0)} \\ \bar{\mathbf{Y}}^{(1)} \end{bmatrix} &= \begin{bmatrix} \Sigma/n_0 & O \\ O & \Sigma/n_1 \end{bmatrix} \end{aligned} .$$

The likelihood ratio test or difference test for $H_0 : \gamma = 0$, based on SEM, is representable as

$$\begin{aligned} &\begin{bmatrix} \bar{\mathbf{Y}}^{(0)} \\ \bar{\mathbf{Y}}^{(1)} \end{bmatrix}' \begin{bmatrix} \Sigma/n_0 & O \\ O & \Sigma/n_1 \end{bmatrix}^{-1} * \\ &\left\{ \begin{bmatrix} I_p \\ I_p \end{bmatrix} \left(\begin{bmatrix} I_p \\ I_p \end{bmatrix}' \begin{bmatrix} \Sigma/n_0 & O \\ O & \Sigma/n_1 \end{bmatrix}^{-1} \begin{bmatrix} I_p \\ I_p \end{bmatrix} \right)^{-1} \begin{bmatrix} I_p \\ I_p \end{bmatrix}' - \right. \\ &\left. \begin{bmatrix} I_p & \mathbf{0} \\ I_p & \boldsymbol{\lambda} \end{bmatrix} \left(\begin{bmatrix} I_p & \mathbf{0} \\ I_p & \boldsymbol{\lambda} \end{bmatrix}' \begin{bmatrix} \Sigma/n_0 & O \\ O & \Sigma/n_1 \end{bmatrix}^{-1} \begin{bmatrix} I_p & \mathbf{0} \\ I_p & \boldsymbol{\lambda} \end{bmatrix} \right)^{-1} \begin{bmatrix} I_p & \mathbf{0} \\ I_p & \boldsymbol{\lambda} \end{bmatrix}' \right\} * \\ &\begin{bmatrix} \Sigma/n_0 & O \\ O & \Sigma/n_1 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{Y}}^{(0)} \\ \bar{\mathbf{Y}}^{(1)} \end{bmatrix} . \end{aligned} \tag{1.17}$$

For details see, *e.g.*, Rao (1973, Chap. 4) for derivation with regard to linear models and/or Browne & Shapiro (1988) for derivation with regard to SEM. The distribution of the statistic above is the chi-square with 1 degree of freedom, and the non-centrality parameter δ^2_{SEM} is obtained by replacing $\begin{bmatrix} \bar{\mathbf{Y}}^{(0)} \\ \bar{\mathbf{Y}}^{(1)} \end{bmatrix}$ with $\begin{bmatrix} \mathbf{0} \\ \boldsymbol{\lambda}\gamma \end{bmatrix}$ in (1.17). After some simplifications, we have

$$\delta^2_{\mathsf{SEM}} = \frac{n_0 n_1 \gamma^2}{n_0 + n_1} \boldsymbol{\lambda}' \Sigma^{-1} \boldsymbol{\lambda} \,. \tag{1.18}$$

Finally, we consider testing the hypothesis using linear combinations of the manifest variables. Note that

$$\begin{aligned} E \begin{bmatrix} \mathbf{c}'\bar{\mathbf{Y}}^{(0)} \\ \mathbf{c}'\bar{\mathbf{Y}}^{(1)} \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 1 & \mathbf{c}'\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \mathbf{c}'\boldsymbol{\nu} \\ \gamma \end{bmatrix} \\ V \begin{bmatrix} \mathbf{c}'\bar{\mathbf{Y}}^{(0)} \\ \mathbf{c}'\bar{\mathbf{Y}}^{(1)} \end{bmatrix} &= \begin{bmatrix} \mathbf{c}'\Sigma\mathbf{c}/n_0 & 0 \\ 0 & \mathbf{c}'\Sigma\mathbf{c}/n_1 \end{bmatrix} . \end{aligned}$$

A similar calculation to (1.17) leads to the chi-square distribution with 1 degree of freedom and the noncentrality parameter as

$$\delta^2_{\mathbf{c}} = \frac{n_0 n_1 \gamma^2}{n_0 + n_1} \cdot \frac{(\mathbf{c}'\boldsymbol{\lambda})^2}{\mathbf{c}'\Sigma\mathbf{c}} \,. \tag{1.19}$$

The Cauchy-Schwarz inequality (see, *e.g.*, Rao 1973, p. 54) shows that

$$\delta^2_{\mathsf{SEM}} \geq \delta^2_c \tag{1.20}$$

and that the equality is attained when

$$\mathbf{c} = \Sigma^{-1}\boldsymbol{\lambda} \,. \tag{1.21}$$

It follows from (1.20) and (1.21) that the SEM approach as well as the factor score approach makes the best inference in terms of statistical power. However, the factor score approach has a drawback when the structural parameters are unknown because replacement of the parameters with their estimates induces dependency between samples, that is, $\hat{\boldsymbol{\lambda}}'\hat{\Sigma}^{-1}\mathbf{Y}_i$'s are not independently distributed. Therefore, the basic assumption in ANOVA is violated.

Finally, we note that the non-centrality parameters in (1.18) and (1.19) are invariant against changing the scale of F_1, *i.e.*, which path coefficient is fixed to be one, because so is $\boldsymbol{\lambda}\gamma$. Another comment is that the quantity $(\mathbf{c}'\boldsymbol{\lambda})^2/(\mathbf{c}'\Sigma\mathbf{c})$ in (1.19) is a reliability coefficient of the weighted scale score $\mathbf{c}'\mathbf{Y}$. The larger the reliability is, the higher is the statistical power of the mean difference test.

1.6 Epilogue

This chapter considers statistical inference via structural equation modeling with independent (exogenous) non-normal, possibly fixed-effect, variables. The introduction of conditional likelihood connects between normal theory inference with such variables and inference by a mean and covariance structure model. The SEM generates a wider class of models than traditional experimental designs. The first extension is to use a model to represent a covariance structure for observed variables $\mathbf{Y}$, not just a saturated structure as in traditional inference. The second is to analyze latent means, not just the difference of general mean vectors. In this chapter, we have just considered the effect of X_1 either on latent means (Fig. 1.4) or on means of $\mathbf{Y}$ (Fig. 1.3). One can also constitute a model in which there is an effect of X_1 on both latent means and means of $\mathbf{Y}$ (see Muthén 1989), so that one can distinguish between a direct mean effect and indirect mean effect of X_1 on $\mathbf{Y}$. The third is to allow for heterogeneous covariance matrices of $\mathbf{Y}$ if one makes multiple population analysis.[4]

The main content of Section 1.4 is nothing but robustness of normal theory inference against non-normal independent variables. Asymptotic robustness study of normal theory inference has been extensively explored by Anderson (1987, 1989), Anderson & Amemiya (1988), Bentler (1983), Browne & Shapiro (1988), Kano (1993), Kano, Berkane, & Bentler (1990), and Satorra (1989), among others. There is close connection between the conditional inference and the asymptotic robustness theory. The conditional inference establishes *exact* robustness whereas the latter shows merely *asymptotic* robustness. But the asymptotic robustness theory allows for non-normal errors.

[4]This topic was skipped.