Chapter 2 Basic Concepts and Applications of Structural Equation Models

The measurement equation in SEMs is a confirmatory tool rather than an exploratory tool, and it can be regarded as a confirmatory factor analysis model. The effects of explanatory latent variables on outcome latent variables are assessed through the structural equation in the model.

Objectives of this chapter:

1. Introduce the basic concepts of SEMs through models with a linear or a nonlinear structural equation.
2. Illustrate how to apply these models to substantive researches.
3. Illustrative the models through real medical studies.

Linear SEMs are formulated with a measurement equation and a linear structural equation. Under the assumption that the observed variables are continuous, and i.i.d. as a normal distribution, the linear SEM is the most basic SEM.

Let $\mathbf{y} = (y_1, \cdots, y_p)^T$ be a $p \times 1$ vector of observed variables that have been selected for the analysis, and let $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_q)^T$ be a $q \times 1$ vector of latent variables that are expected to be formed from the observed variables in $\mathbf{y}$. The link between the observed variables and all the latent variables in $\boldsymbol{\omega}$ is defined by the following measurement equation: For $j = 1, \cdots, p$,

$$y_j = \mu_j + \lambda_{j1}\omega_1 + \cdots + \lambda_{jq}\omega_q + \epsilon_j, \tag{1}$$

where μ_j is an intercept, λ_{jk}'s are unknown coefficients that relate y_j and ω_k, and ϵ_j is the residual error. In the factor analysis terminology, λ_{jk}'s are called factor loadings.

Let $\boldsymbol{\omega} = (\boldsymbol{\eta}^T, \boldsymbol{\xi}^T)^T$, where $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are $q_1 \times 1$ and $q_2 (= q - q_1) \times 1$ random vectors which respectively contain the outcome and explanatory latent variables in $\boldsymbol{\omega}$. The effects of $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_{q_2})^T$ on $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_{q_1})^T$ are assessed by the following structural equation: For $j = 1, \cdots, q_1$,

$$\eta_j = \gamma_{j1}\xi_1 + \cdots + \gamma_{jq_2}\xi_{q_2} + \delta_j, \tag{2}$$

where γ_{jk}'s are unknown coefficients that represent the effects of ξ_k on η_j, and δ_j is the residual error. Equations (1) and (2) define the most basic linear SEM. They can be rewritten in matrix notation:

$$\text{Measurement Equation : } \mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \tag{3}$$

$$\text{Structural Equation : } \boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{4}$$

where $\mathbf{y}$ is a $p \times 1$ random vector of observed variables, $\boldsymbol{\mu}$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times q$ unknown matrix of factor loadings, $\boldsymbol{\Gamma}$ is a $q_1 \times q_2$ unknown matrix of regression coefficients, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are $p \times 1$ and $q_1 \times 1$ random vectors of measurement (residual) errors, respectively.

The purpose of the measurement equation in an SEM is to relate the latent variables in $\boldsymbol{\omega}$ to the observed variables in $\mathbf{y}$. It represents the link between observed and latent variables, through the specified factor loading matrix $\boldsymbol{\Lambda}$. The vector of measurement error, $\boldsymbol{\epsilon}$, is used for taking the residual errors into account.

The most important issue in formulating the measurement equation is to specify the structure of the factor loading matrix, $\boldsymbol{\Lambda}$, based on the knowledge of the observed variables in the study. Any element of $\boldsymbol{\Lambda}$ can be a free parameter or a fixed parameter with a preassigned value. The positions and the preassigned values of fixed parameters are decided on the basis of the prior knowledge of the observed variables and latent variables, and they are also related to the interpretations of latent variables. We give a simple example to illustrate the formulation.

Consider a study concerning the effects of blood pressure and obesity on kidney disease of type 2 diabetic patients. From its objective, we are interested in three latent variables, namely one outcome latent variable about kidney disease, and two explanatory latent variables about blood pressure and obesity. Based on the related medical knowledge,

- plasma creatine (PCr) and urinary albumin creatinine ratio (ACR) are grouped into 'kidney disease (KD)';
- systolic blood pressure (SBP) and diastolic blood pressure (DBP) are grouped into 'blood pressure (BP)';
- body mass index (BMI) and waist hip ratio (WHR) are grouped into 'obesity (OB)'.

From clear interpretation of BP, and the meaning of the observed variables, BP should only relate to SBP and DBP, but not to other observed variables. This rationale also applies to latent variables KD and OB.

Thus, the system of measurement equations is defined as:

$$\begin{aligned}
\text{PCr} &= \mu_1 + \lambda_{11}\text{KD} + \epsilon_1, \\
\text{ACR} &= \mu_2 + \lambda_{21}\text{KD} + \epsilon_2, \\
\text{SBP} &= \mu_3 + \lambda_{32}\text{BP} + \epsilon_3, \\
\text{DBP} &= \mu_4 + \lambda_{42}\text{BP} + \epsilon_4, \\
\text{BMI} &= \mu_5 + \lambda_{53}\text{OB} + \epsilon_5, \\
\text{WHR} &= \mu_6 + \lambda_{63}\text{OB} + \epsilon_6,
\end{aligned} \tag{5}$$

or in matrix notation:

$$\begin{bmatrix} \text{PCr} \\ \text{ACR} \\ \text{SBP} \\ \text{DBP} \\ \text{BMI} \\ \text{WHR} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & \lambda_{53} \\ 0 & 0 & \lambda_{63} \end{bmatrix} \begin{bmatrix} \text{KD} \\ \text{BP} \\ \text{OB} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}, \tag{6}$$

or

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\omega}$, and $\boldsymbol{\epsilon}$ are defined as in (3). From the structure of $\boldsymbol{\Lambda}$, we know that KD is only linked with PCr and ACR, BP is only linked with SBP and DBP, and OB is only linked with BMI and WHR. As a result, the interpretation of the latent variables, KD, BP, and OB, is clear. This specific structure of $\boldsymbol{\Lambda}$ is called a non-overlapping structure.

In most situations, it is not necessary to use a more general structure of $\boldsymbol{\Lambda}$. For example, if λ_{12} in Equation (6) is nonzero, then BP is also related to PCr. Hence, BP cannot be interpreted as blood pressure, and the effect of blood pressure on kidney disease cannot be clearly assessed. In this book, we use $\boldsymbol{\Lambda}$ with a non-overlapping structure in all real applications.

Recall that $\boldsymbol{\omega} = (\boldsymbol{\eta}^T, \boldsymbol{\xi}^T)^T$. The choices of the outcome and explanatory latent variables are based on the objective of the substantive study. Subsequently, q_1 and q_2 are defined.

In the kidney disease study, KD is the outcome latent variable, and BP and OB are the explanatory latent variables; hence, $q_1 = 1$ and $q_2 = 2$. The structural equation can be defined as:

$$\mathsf{KD} = \gamma_1 \mathsf{BP} + \gamma_2 \mathsf{OB} + \delta. \tag{7}$$

This equation is linear in the variables and linear in the parameters. The interpretations of γ_1 and γ_2 are the same as in a regression model. Hence, they represent the magnitude of the expected changes in KD for one unit change in BP and OB, respectively. The outcome latent variables are only partially explained by the explanatory latent variables, the unexplained part is taken into account by the residual error δ.

A slight extension of the structural equation (4) that is particularly useful in business and social-psychological research is defined by

$$\boldsymbol{\eta} = \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{8}$$

where $\boldsymbol{\Pi}$ is a $q_1 \times q_1$ matrix of unknown coefficients, such that $\mathbf{I} - \boldsymbol{\Pi}$ is nonsingular and the diagonal elements of $\boldsymbol{\Pi}$ are zero; and the definitions of $\boldsymbol{\Gamma}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ are the same as before. According to specific applications, elements in $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}$ can be fixed to preassigned values.

This structural equation allows some outcome latent variables depend on the other outcome latent variables through an appropriately defined $\boldsymbol{\Pi}$.

For example, we wish to study the effects of BP and OB on KD, as well as a disease A, η_A. Suppose that it is also interesting to examine the possible effect of KD on disease A. The following structural equation can be used:

$$\begin{pmatrix} \text{KD} \\ \eta_A \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \pi & 0 \end{pmatrix} \begin{pmatrix} \text{KD} \\ \eta_A \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{pmatrix} \begin{pmatrix} \text{BP} \\ \text{OB} \end{pmatrix} + \begin{pmatrix} \delta \\ \delta_A \end{pmatrix}. \tag{9}$$

Here, $\boldsymbol{\eta} = (\text{KD}, \eta_A)^T$, π is the unknown coefficient that represents the effect of KD on disease A, and $\boldsymbol{\Gamma}$ is the parameter matrix with elements γ_i. Equation (9) can be rewritten as

$$\begin{aligned} \text{KD} &= \gamma_1 \text{BP} + \gamma_2 \text{OB} + \delta, \\ \eta_A &= \pi \text{KD} + \gamma_3 \text{BP} + \gamma_4 \text{OB} + \delta_A. \end{aligned} \tag{10}$$

By allowing elements in $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}$ to be fixed at any preassigned values, structural equation (8) achieves considerable flexibility in handling rather complex relationships among latent variables.

The standard linear SEMs have some assumptions: For $i = 1, \cdots, n$,

A1: The random vectors of residual errors $\boldsymbol{\epsilon}_i$ are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, where $\boldsymbol{\Psi}_\epsilon$ is a diagonal covariance matrix.

A2: The random vectors of explanatory latent variables $\boldsymbol{\xi}_i$ are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$, where $\boldsymbol{\Phi}$ is a general covariance matrix.

A3: The random vectors of residual errors $\boldsymbol{\delta}_i$ are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$, where $\boldsymbol{\Psi}_\delta$ is a diagonal covariance matrix.

A4: $\boldsymbol{\delta}_i$ is independent of $\boldsymbol{\xi}_i$, and $\boldsymbol{\epsilon}_i$ is independent of $\boldsymbol{\omega}_i$ and $\boldsymbol{\delta}_i$.

These assumptions imply that

- $\boldsymbol{\eta}_i$ are i.i.d. normal because $\boldsymbol{\xi}_i$ and $\boldsymbol{\delta}_i$ are i.i.d. normal;
- $\boldsymbol{\omega}_i$ are i.i.d. normal because $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$ are i.i.d. normal;
- $\mathbf{y}_i$ are i.i.d. normal because $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ are i.i.d. normal.

Model identification is an issue relevant to all SEMs. Let $\boldsymbol{\theta}$ be a parameter vector. The traditional definition of identification is based on $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the population covariance matrix of the observed variables in $\mathbf{y}$. The model is said to be identified if for any $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ (Bollen, 1989).

This definition is difficult to apply to a complex SEM whose $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is very complicated or even impossible to derive. Hence, we consider a definition of identification on the basis of the measurement equation $m(\boldsymbol{\theta})$ and structural equation $s(\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ have no common element. Specifically,

- The measurement equation is identified if for any $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, $m(\boldsymbol{\theta}_1) = m(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$;
- The structural equation is identified if for any $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$, $s(\boldsymbol{\theta}_1^*) = s(\boldsymbol{\theta}_2^*)$ implies $\boldsymbol{\theta}_1^* = \boldsymbol{\theta}_2^*$;
- The SEM is identified if both of its measurement equation and structural equation are identified.

General necessary and sufficient conditions to guarantee the identifiability of an SEM are difficult to find. Hence, in practical applications of SEMs, we mainly concern the sufficient conditions for achieving an identified model.

Consider linear SEMs. The measurement equation is not identified without imposing some identification condition. For any nonsingular matrix $\mathbf{M}$,

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{M}\mathbf{M}^{-1}\boldsymbol{\omega} + \boldsymbol{\epsilon} \tag{11}$$

$$= \boldsymbol{\mu} + \boldsymbol{\Lambda}^{*}\boldsymbol{\omega}^{*} + \boldsymbol{\epsilon}, \tag{12}$$

where $\boldsymbol{\Lambda}^{*} = \boldsymbol{\Lambda}\mathbf{M}$, and $\boldsymbol{\omega}^{*} = \mathbf{M}^{-1}\boldsymbol{\omega}$, which is a random vector of latent variables with distribution $N[\mathbf{0}, \mathbf{M}^{-1}\boldsymbol{\Phi}^{+}(\mathbf{M}^{-1})^{T}]$, where $\boldsymbol{\Phi}^{+}$ is the covariance matrix of $\boldsymbol{\omega}$. To identify the measurement equation, we have to impose restrictions on $\boldsymbol{\Lambda}$ and/or $\boldsymbol{\Phi}^{+}$, such that the only nonsingular matrix $\mathbf{M}$ that satisfies the imposed conditions is the identity matrix.

A simple and common method is using a $\boldsymbol{\Lambda}$ with the non-overlapping structure.

Consider an example with $p = 10$ and $q = 3$, in which the first four observed variables are related to ω_1, the next and the last three observed variables are related to ω_2 and ω_3, respectively. A non-overlapping structure of $\mathbf{\Lambda}$ is given as follows:

$$\mathbf{\Lambda}^T = \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & \lambda_{41} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{62} & \lambda_{72} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{93} & \lambda_{10,3} \end{bmatrix},$$

where 1's and 0's are fixed, and the λ_{jk}'s are unknown parameters. Note that

1. The fixed value 1 is used to introduce a scale to latent variables.
2. The choice of λ_{11}, λ_{52}, and λ_{83} is only for convenience.
3. Based on the meaning of the observed variables, we have a clear idea about the positions of the parameters fixed as 0.

There are other methods to identify the measurement equation. For instance, we may allow λ_{11}, λ_{52}, and/or λ_{83} in $\mathbf{\Lambda}$ to be unknown parameters, and fix the diagonal elements of $\mathbf{\Phi}^+$ as 1. This method restricts the variances of latent variables to be 1; hence $\mathbf{\Phi}^+$ is a correlation matrix. As this method is not convenient for identifying an SEM with a structural equation, and it induces complication in the Bayesian analysis (see Chapter 3), we use the first identification method to identify the measureemnt equation throughout this book.

For almost all applications of SEMs, the structural equation is identified with identified $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. If necessary, the above simple method (via fixing appropriate parameters) for identifying the measurement equation can be used to identify the structural equation.

A path diagram is a pictorial representation of the measurement and structural equations. The following conventions are assumed:

(I) Observed variables such as x- and y-variables are enclosed in rectangles or squares. Latent variables such as ξ- and η-variables are enclosed in ellipses or circles. Residual errors such as δ and ϵ are included in the path diagram but are not enclosed.

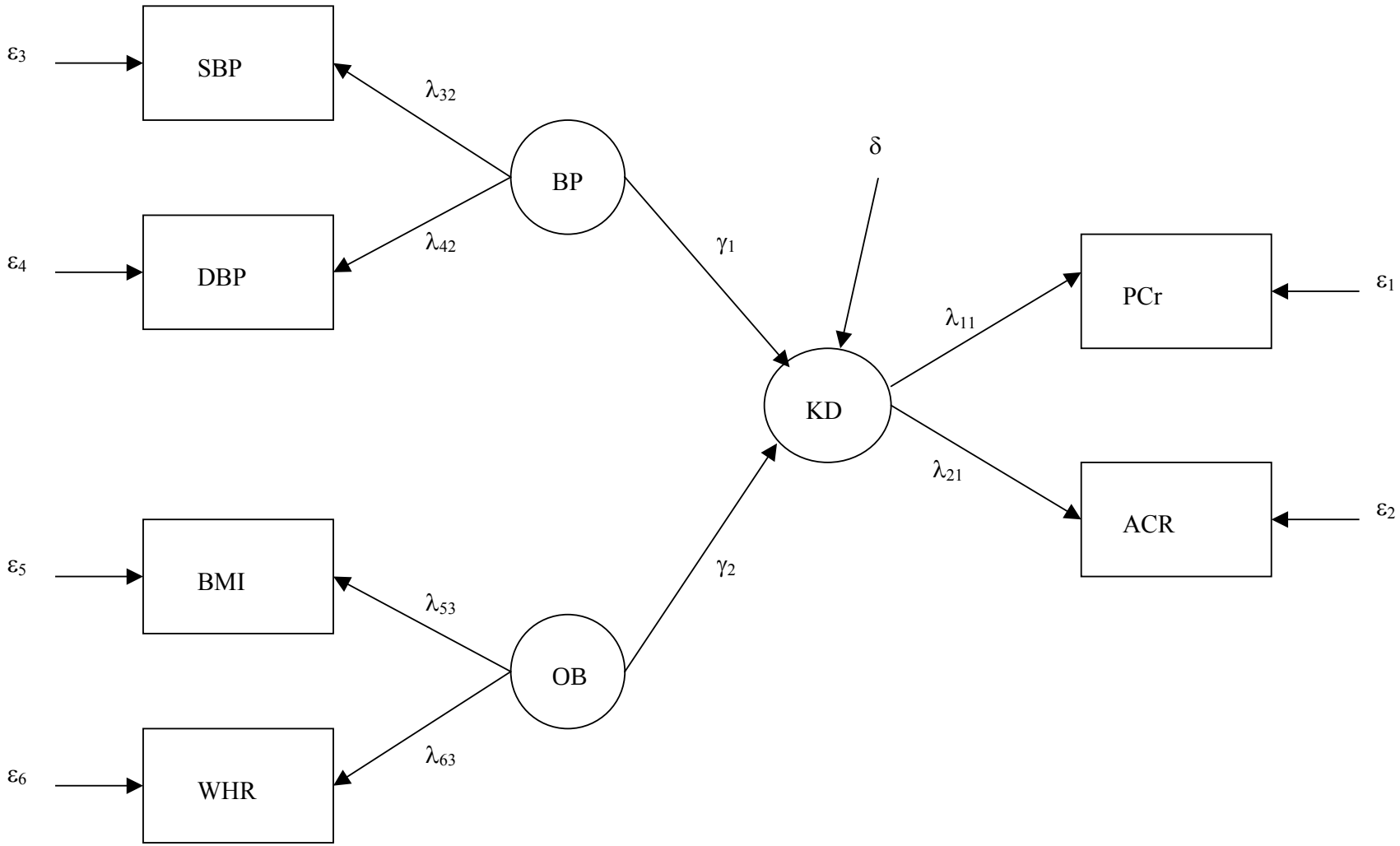
(II) A one-way arrow between two variables indicates a postulated direct influence of one variable on another. A two-way arrow between two variables indicates that these variables may be correlated.

(III) The coefficient associated with each arrow indicates the corresponding parameter.

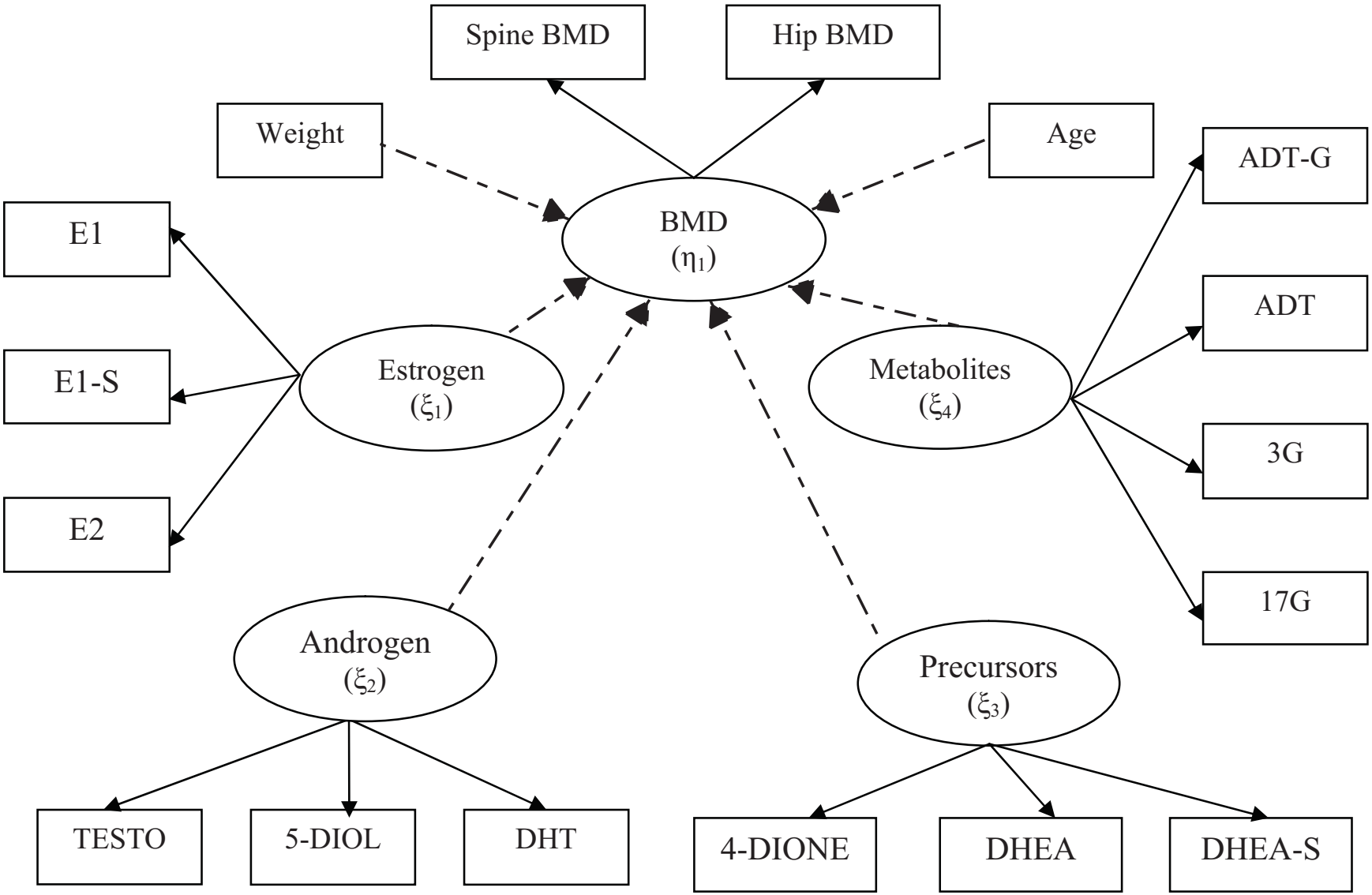
(IV) All direct influences of one variable on another are included in the path diagram.

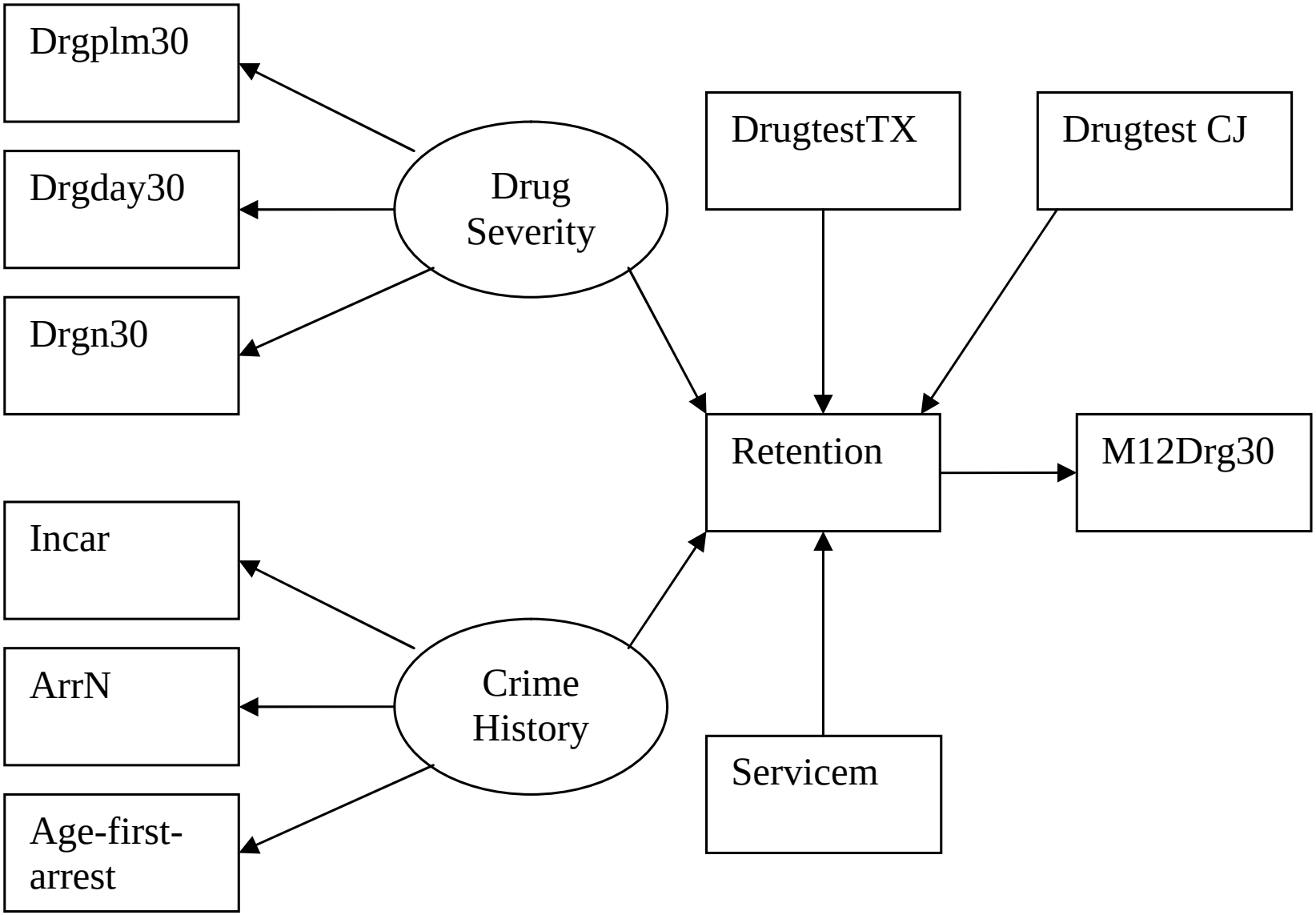
Sometimes, two-way arrows between two correlated variables and/or residual errors are not drawn for clarity. Moreover, the means (intercepts) may not be presented in the path diagram.

The path diagram related to the SEM with measurement equation (6) and structural equation (7) is presented in Figure 2.1.



Other examples:




Drgplm30
Drgday30
Drgn30
Drug Severity
DrugtestTX
Drugtest CJ
Retention
M12Drg30
Incar
ArrN
Age-first-arrest
Crime History
Servicem

To develop better models, it is often desirable to incorporate explanatory observed variables on the right-hand sides of the measurement and structural equations. In the field of SEM, these explanatory observed variables are regarded as fixed covariates.

Accommodation of fixed covariates in the measurement equation provides additional information about the latent exposure and thus reduces estimation uncertainty for the latent variables. For the structural equation, fixed covariates give more ingredients to account for the outcome latent variables, in addition to the explanatory latent variables. Hence, the residual errors in both equations can be reduced by incorporating fixed covariates.

SEMs with fixed covariates are defined as follows. The measurement equation is given by:

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \tag{13}$$

where $\mathbf{A}$ is a $p \times r_1$ matrix of unknown coefficients, $\mathbf{c}$ is an $r_1 \times 1$ vector of fixed covariates, and $\boldsymbol{\Lambda}$, $\boldsymbol{\omega}$, and $\boldsymbol{\epsilon}$ are defined as before. A simple example with one intercept and one fixed covariate c_2 is:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ \vdots & \vdots \\ a_{p1} & a_{p2} \end{bmatrix} \begin{bmatrix} 1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1q} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pq} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_q \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix}, \tag{14}$$

or

$$y_j = a_{j1} + a_{j2}c_2 + \lambda_{j1}\omega_1 + \cdots + \lambda_{jq}\omega_q + \epsilon_j, \quad j = 1, \cdots, p.$$

If $c_2 = 0$, and let $\mu_j = a_{j1}$, Equation (14) reduces to Equation (3).

The structural equation is defined by

$$\boldsymbol{\eta} = \mathbf{B}\mathbf{d} + \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{15}$$

where $\mathbf{B}$ is a $q_1 \times r_2$ matrix of unknown coefficients, $\mathbf{d}$ is an $r_2 \times 1$ vector of fixed covariates, and $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\delta}$ are defined as before. Note that $\mathbf{c}$ and $\mathbf{d}$ may have common elements; and (15) reduces to (8) if $\mathbf{d} = \mathbf{0}$. A simple example is

$$\eta = b_1 d_1 + b_2 d_2 + \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \delta,$$

where $\mathbf{B} = (b_1, b_2)$, and $\boldsymbol{\Gamma} = (\gamma_1, \gamma_2, \gamma_3)$.

The assumptions of SEMs with fixed covariates are the same as A1, A2, A3, and A4. As fixed covariates are observed, the distributions of $\boldsymbol{\omega}_i$ and $\mathbf{y}_i$ are still normal. Similar to the basic linear SEMs, SEMs with fixed covariates can be identified by fixing appropriate parameters at given values.

Suppose that the main objective is on studying the complex diabetic kidney disease, with emphasis on assessing effects of blood pressure, obesity, lipid control as well as some covariates on that disease. Based on some known medical knowledge, data related to the observed variables $(y_1, \cdots, y_9)$ were collected to form latent variables $(\eta, \xi_1, \cdots, \xi_3)$ as follows:

- {PCr, ACR} — kidney disease (KD)
- {SBP, DBP} — blood pressure (BP)
- {BMI, WHR} — obesity (OB)
- {non-high-density lipoprotein cholesterol (non-HDL-C), low-density lipoprotein cholesterol (LDL-C), plasma triglyceride (TG)} — lipid control (LIP)

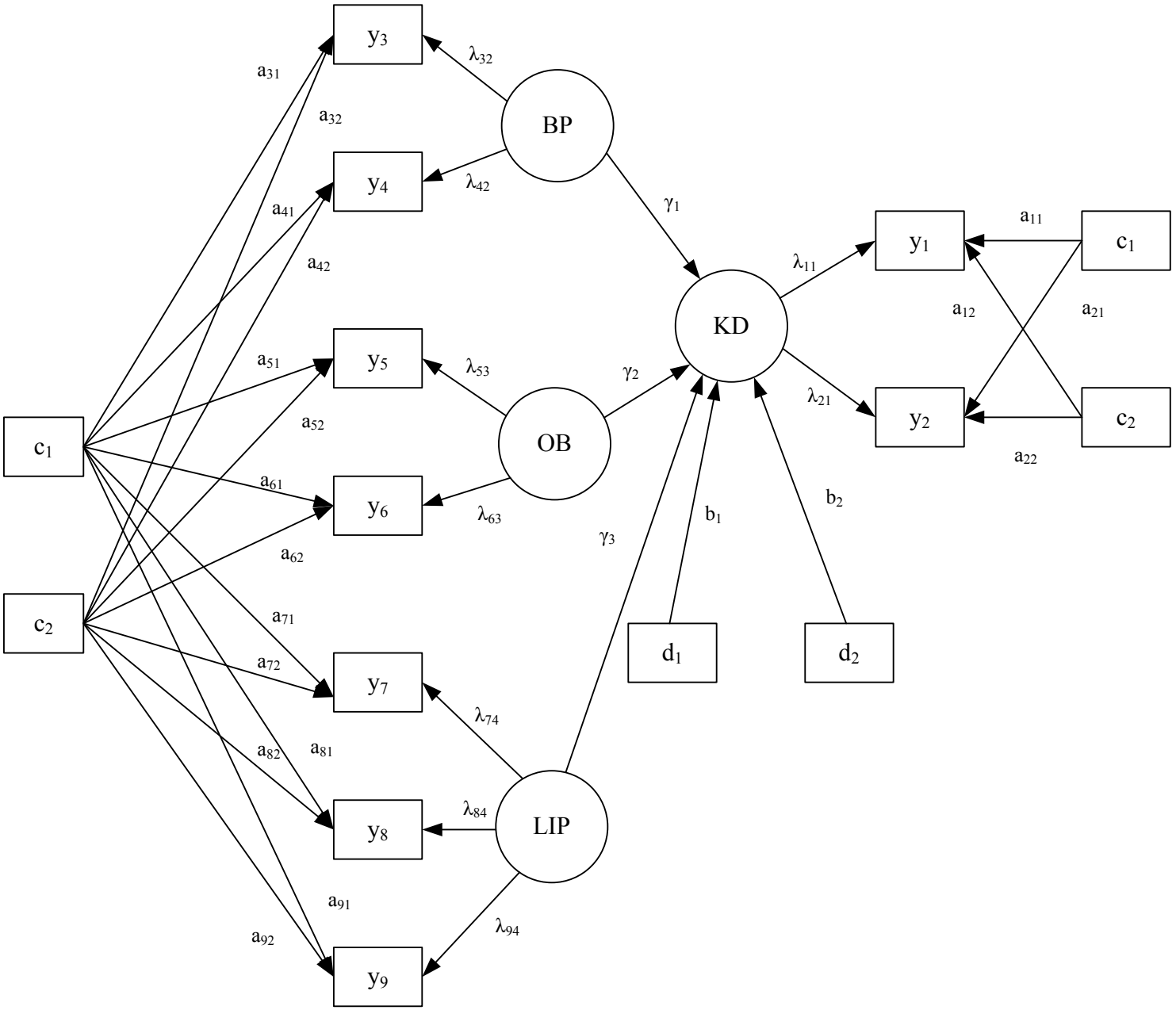
Now, we incorporate 'smoking (c_1)' and 'alcohol (c_2)' in the measurement equation, and 'age (d_1)' and 'gender (d_2)' in the structural equation.

The SEM that incorporates fixed covariates c_1 and c_2 in the measurement equation, and d_1 and d_2 in the structural equation are given below:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \\ a_{71} & a_{72} \\ a_{81} & a_{82} \\ a_{91} & a_{92} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ 0 & \lambda_{32} & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 \\ 0 & 0 & \lambda_{53} & 0 \\ 0 & 0 & \lambda_{63} & 0 \\ 0 & 0 & 0 & \lambda_{74} \\ 0 & 0 & 0 & \lambda_{84} \\ 0 & 0 & 0 & \lambda_{94} \end{bmatrix} \begin{bmatrix} \mathsf{KD} \\ \mathsf{BP} \\ \mathsf{OB} \\ \mathsf{LIP} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix}. \quad (16)$$

$$\mathsf{KD} = b_1 \mathsf{age} + b_2 \mathsf{gender} + \gamma_1 \mathsf{BP} + \gamma_2 \mathsf{OB} + \gamma_3 \mathsf{LIP} + \delta, \quad (17)$$

where a_{jk}, λ_{jk}, b_1, b_2, γ_1, γ_2, and γ_3 are unknown regression coefficients.


y3
y4
y5
y6
y7
y8
y9
y1
y2
BP
OB
LIP
KD
c1
c2
d1
d2
λ32
λ42
λ53
λ63
λ74
λ84
λ94
λ11
λ21
γ1
γ2
γ3
b1
b2
a31
a32
a41
a42
a51
a52
a61
a62
a71
a72
a81
a82
a91
a92
a11
a12
a21
a22

The measurement equation of nonlinear SEMs is defined as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \tag{18}$$

which has exactly the same form as in (3). The structural equation is formulated as

$$\boldsymbol{\eta} = \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}) + \boldsymbol{\delta}, \tag{19}$$

where $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\delta}$ are similarly defined as before, $\mathbf{F}(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), \cdots, f_t(\boldsymbol{\xi}))^T$ is a $t \times 1$ vector-valued function with nonzero, known, and linearly independent differentiable functions $f_1, \cdots, f_t$, and $t \geq q_2$.

The assumptions of this model are the same as A1, A2, A3, and A4. Due to the presence of the nonlinear terms of $\boldsymbol{\xi}$ in $\mathbf{F}(\boldsymbol{\xi})$, the distributions of $\boldsymbol{\omega}_i$ and $\mathbf{y}_i$ are no longer normal. In other words, nonlinear SEMs do not assume that $\boldsymbol{\omega}_i$ and $\mathbf{y}_i$ are normal.

Similar to linear SEMs, the measurement equation of nonlinear SEMs can be identified by fixing appropriate parameters at some given values.

To achieve an identified structural equation, the choice of $\mathbf{F}(\boldsymbol{\xi})$ in the structural equation is not completely arbitrary. For example, the following obvious cases are not allowed: $\mathbf{F}_1(\boldsymbol{\xi}) = (\xi_1, \xi_2, \xi_1^2, \xi_1^2)^T$ and $\mathbf{F}_2(\boldsymbol{\xi}) = (\xi_1, \xi_2, \xi_1\xi_2, 0)^T$. They should be modified as $\mathbf{F}_1(\boldsymbol{\xi}) = (\xi_1, \xi_2, \xi_1^2)^T$ and $\mathbf{F}_2(\boldsymbol{\xi}) = (\xi_1, \xi_2, \xi_1\xi_2)^T$, respectively. An example of identified structural equations is :

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & \pi \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1^2 \\ \xi_1\xi_2 \\ \xi_2^2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}.$$

More care is needed to interpret the mean vector of $\mathbf{y}$. Let $\mathbf{\Lambda}_k^T$ be the kth row of $\mathbf{\Lambda}$. For $k = 1, \cdots, p$, it follows from Equation (3) that

$$E(y_k) = \mu_k + \mathbf{\Lambda}_k^T E(\boldsymbol{\omega}).$$

Although $E(\boldsymbol{\xi}) = \mathbf{0}$, it follows from $\boldsymbol{\eta} = \mathbf{\Pi}\boldsymbol{\eta} + \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}) + \boldsymbol{\delta}$ that $E(\boldsymbol{\eta}) \neq \mathbf{0}$ if $\mathbf{F}(\boldsymbol{\xi})$ is a vector-valued nonlinear function of $\boldsymbol{\xi}$ and $E(\mathbf{F}(\boldsymbol{\xi})) \neq 0$. Hence $E(\boldsymbol{\omega}) \neq \mathbf{0}$ and $E(y_k) \neq \mu_k$.

Let $\mathbf{\Lambda}_k^T = (\mathbf{\Lambda}_{k\eta}^T, \mathbf{\Lambda}_{k\xi}^T)$ be a partition of $\mathbf{\Lambda}_k^T$ that corresponds to the partition of $\boldsymbol{\omega} = (\boldsymbol{\eta}^T, \boldsymbol{\xi}^T)^T$. Because $E(\boldsymbol{\xi}) = \mathbf{0}$ and $E(\boldsymbol{\eta}) = [(\mathbf{I} - \mathbf{\Pi})^{-1}\mathbf{\Gamma}]E(\mathbf{F}(\boldsymbol{\xi}))$,

$$E(y_k) = \mu_k + \mathbf{\Lambda}_{k\eta}^T E(\boldsymbol{\eta}) + \mathbf{\Lambda}_{k\xi}^T E(\boldsymbol{\xi}) = \mu_k + \mathbf{\Lambda}_{k\eta}^T [(\mathbf{I} - \mathbf{\Pi})^{-1}\mathbf{\Gamma}]E(\mathbf{F}(\boldsymbol{\xi})). \quad (20)$$

Linear SEMs with fixed covariates can be naturally generalized to nonlinear SEMs with fixed covariates through the following measurement and structural equations:

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \tag{21}$$

$$\boldsymbol{\eta} = \mathbf{B}\mathbf{d} + \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}) + \boldsymbol{\delta}, \tag{22}$$

where the definitions of the random vectors and the parameter matrices are the same as before.

In this model, the measurement equation is the same as in (13), while the structural equation can be regarded as a natural extension of equations (15) and (19).

As a simple example, we consider a continuation of the previous artificial example. Suppose that we wish to study various interactive effects of the explanatory latent variables BP, OB, and LIP on KD. To achieve our goal, we consider a model with its measurement equation given in (16), while the structural equation is formulated as

$$\begin{aligned} \text{KD} =& b_1 d_1 + b_2 d_2 + \gamma_1 \text{BP} + \gamma_2 \text{OB} + \gamma_3 \text{LIP} + \\ & \gamma_4 (\text{BP} \times \text{OB}) + \gamma_5 (\text{BP} \times \text{LIP}) + \gamma_6 (\text{OB} \times \text{LIP}) + \delta. \end{aligned} \tag{23}$$

In this formulation, $\mathbf{B} = (b_1, b_2)$, $\mathbf{d} = (d_1, d_2)^T$, $\mathbf{\Gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6)$, and $\mathbf{F}(\boldsymbol{\xi}) = (\text{BP}, \text{OB}, \text{LIP}, \text{BP} \times \text{OB}, \text{BP} \times \text{LIP}, \text{OB} \times \text{LIP})^T$.

A simple extension of the structural equation is:

$$\boldsymbol{\eta} = \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Lambda}_\omega \mathbf{G}(\mathbf{d}, \boldsymbol{\xi}) + \boldsymbol{\delta}. \tag{24}$$

where $\mathbf{G}(\mathbf{d}, \boldsymbol{\xi}) = (g_1(\mathbf{d}, \boldsymbol{\xi}), \cdots, g_t(\mathbf{d}, \boldsymbol{\xi}))^T$ is a vector-valued function with nonzero, known, and linearly independent differentiable functions. A special case of this general structural equation is the one defined by (24) with $\boldsymbol{\Lambda}_\omega = (\mathbf{B}, \boldsymbol{\Gamma})$ and $\mathbf{G}(\mathbf{d}, \boldsymbol{\xi}) = (\mathbf{d}^T, \mathbf{F}(\boldsymbol{\xi})^T)^T$.

Care should be taken to interpret the mean of $\mathbf{y}$. Using the same notation, we have

$$E(y_k) = \mathbf{A}_k^T \mathbf{c} + \boldsymbol{\Lambda}_{k\eta}^T [(\mathbf{I} - \boldsymbol{\Pi})^{-1} \boldsymbol{\Lambda}_\omega] E[\mathbf{G}(\mathbf{d}, \boldsymbol{\xi})], \tag{25}$$

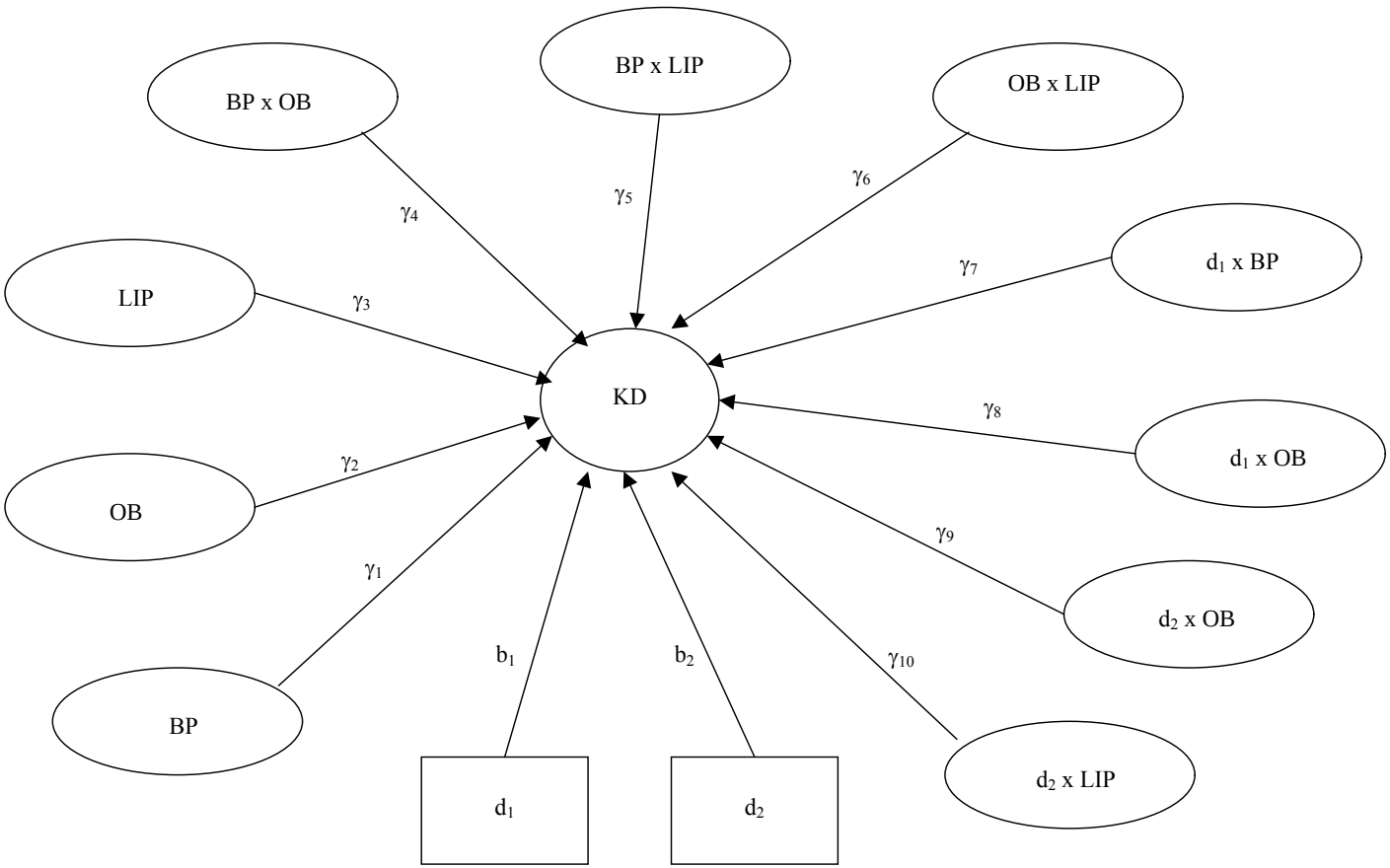
where $\mathbf{A}_k^T$ is the kth row of $\mathbf{A}$.

The artificial example presented above is used again to illustrate the key idea of incorporating nonlinear terms of fixed covariates and explanatory latent variables in the structural equation. While the measurement equation is defined by (16), and the structural equation can be formulated as follows:

$$\begin{aligned} \mathrm{KD} &= b_1 d_1 + b_2 d_2 + \gamma_1 \mathrm{BP} + \gamma_2 \mathrm{OB} + \gamma_3 \mathrm{LIP} + \gamma_4 (\mathrm{BP} \times \mathrm{OB}) \\ &+ \gamma_5 (\mathrm{BP} \times \mathrm{LIP}) + \gamma_6 (\mathrm{OB} \times \mathrm{LIP}) + \gamma_7 (d_1 \times \mathrm{BP}) \\ &+ \gamma_8 (d_1 \times \mathrm{OB}) + \gamma_9 (d_2 \times \mathrm{OB}) + \gamma_{10} (d_2 \times \mathrm{LIP}) + \delta. \end{aligned} \tag{26}$$

Note that more complex product terms of d_1, d_2, BP, OB, and LIP can be assessed via other appropriately defined structural equations.

The path diagram corresponding to the structural equation (26) is presented in Figure 2.3.



Although the emphasis of SEMs is on assessing the effects of explanatory latent variables on the key outcome latent variables, some particular explanatory latent variables may be influenced by other explanatory latent variables and fixed covariates.

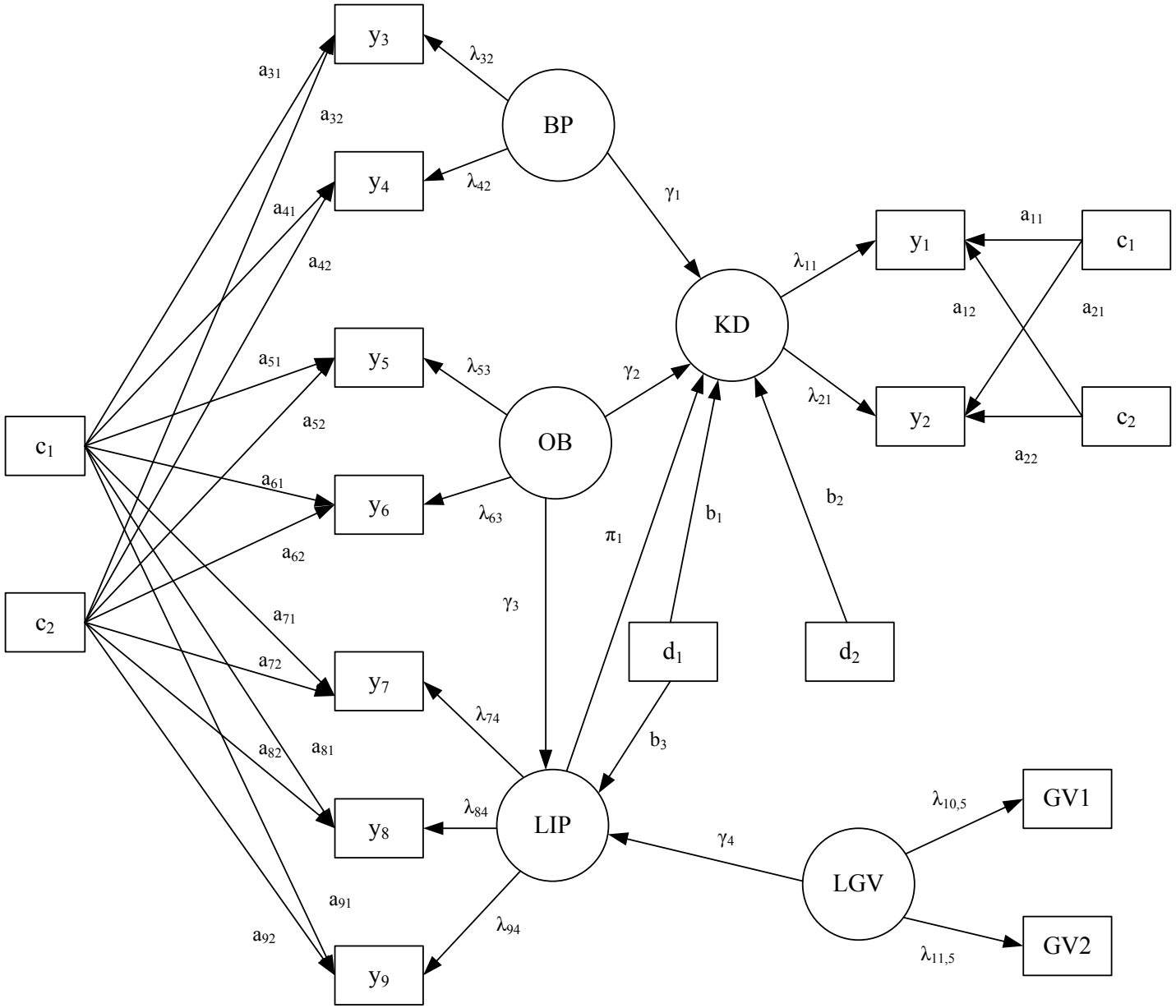
For instance, in the previous artificial example, although KD is the key outcome latent variable, BP, OB, d_1 (age), and d_2 (gender) are also expected to influence the latent variable LIP. To provide a more concrete illustration, suppose that we are interested in the SEM presented in Figure 2.4. Compared to the SEM presented in Figure 2.2, we further have:

(1) two observed genetic variables GV1 and GV2 which form a latent variable LGV,

(2) a path from LGV to LIP,

(3) a path from OB to LIP,

(4) a path from age (b_3) to LIP.


y3
y4
y5
y6
y7
y8
y9
BP
OB
LIP
KD
LGV
y1
y2
c1
c2
d1
d2
GV1
GV2
λ32
λ42
λ53
λ63
λ74
λ84
λ94
λ11
λ21
λ10,5
λ11,5
γ1
γ2
γ3
γ4
π1
b1
b2
b3
a11
a12
a21
a22
a31
a32
a41
a42
a51
a52
a61
a62
a71
a72
a81
a82
a91
a92

The measurement equation is defined by:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ GV_1 \\ GV_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \\ a_{71} & a_{72} \\ a_{81} & a_{82} \\ a_{91} & a_{92} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{32} & 0 & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{53} & 0 & 0 \\ 0 & 0 & \lambda_{63} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{74} & 0 \\ 0 & 0 & 0 & \lambda_{84} & 0 \\ 0 & 0 & 0 & \lambda_{94} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{10,5} \\ 0 & 0 & 0 & 0 & \lambda_{11,5} \end{bmatrix} \begin{bmatrix} KD \\ BP \\ OB \\ LIP \\ LGV \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \end{bmatrix}. \tag{27}$$

To formulate the structural equation associated with this path diagram, LIP is treated as an outcome latent variable. The structural equation is defined as:

$$\begin{pmatrix} \text{KD} \\ \text{LIP} \end{pmatrix} = \begin{pmatrix} b_1 & b_2 \\ b_3 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} + \begin{pmatrix} 0 & \pi_1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \text{KD} \\ \text{LIP} \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_2 & 0 \\ 0 & \gamma_3 & \gamma_4 \end{pmatrix} \begin{pmatrix} \text{BP} \\ \text{OB} \\ \text{LGV} \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}. \tag{28}$$

Here, as OB has a direct effect on LIP, which itself also has a direct effect on KD, OB has an indirect effect ($\pi_1 \times \gamma_3$) on KD. Hence, the total effect of OB on KD is $\gamma_2 + \pi_1\gamma_3$. The above modeling concerning LIP can be considered for explanatory latent variables BP and OB.

There is a temptation to develop a comprehensive SEM that takes into account all the interrelationships of the observed and latent variables. From a practical point of view, the following issues have to be considered:

(I) It is important to make sure the sample size of the available data set is large enough to achieve accurate statistical results.

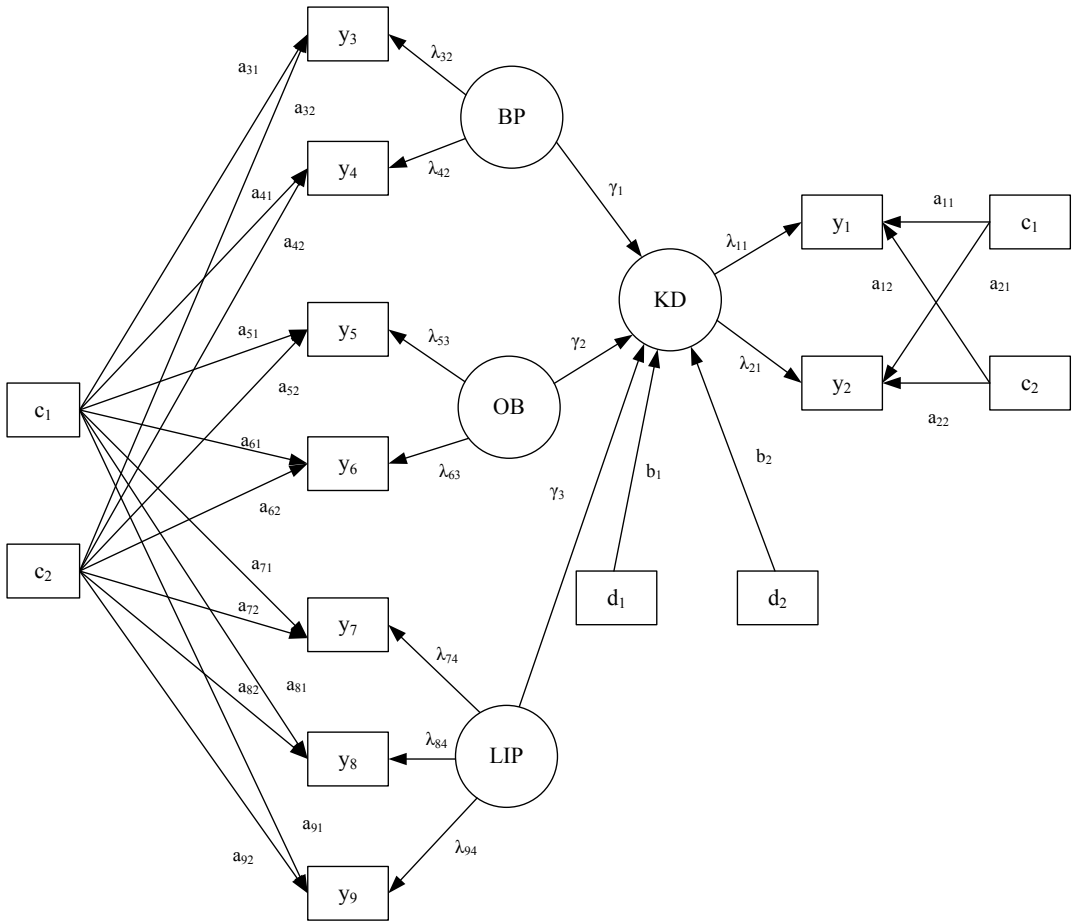
(II) If the size of the proposed SEM and the number of parameters are large, we may encounter difficulties in achieving convergence of the related computing algorithm for obtaining statistical results.

(III) In this chapter, the most general SEM is

$$\mathbf{y} = \mathbf{Ac} + \boldsymbol{\Lambda\omega} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\eta} = \boldsymbol{\Pi\eta} + \boldsymbol{\Lambda}_{\omega}\mathbf{G}(\mathbf{d}, \boldsymbol{\xi}) + \boldsymbol{\delta}.$$

This model has limitations. As $\mathbf{G}(\mathbf{d}, \boldsymbol{\xi})$ does not involve any outcome latent variables in $\boldsymbol{\eta}$, once a latent variable is treated as an outcome latent variable, nonlinear terms related to this latent variable cannot be used to predict the other outcome variables. e.g., as LIP in (28) is treated as an outcome variable, it cannot be accommodated in $\mathbf{G}(\mathbf{d}, \boldsymbol{\xi})$, and nonlinear effects of LIP on KD cannot be assessed.

Under complex situations, we sometimes separate the comprehensive model into submodels. For example, the comprehensive SEM represented in the following path diagram can be separated into two submodels. One is the SEM represented below with main focus on KD.



Here, LIP is only treated as an explanatory latent variable. Its nonlinear effects BP$\times$LIP, OB$\times$LIP, and $d_2\times$LIP on KD can be assessed.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \\ a_{71} & a_{72} \\ a_{81} & a_{82} \\ a_{91} & a_{92} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ 0 & \lambda_{32} & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 \\ 0 & 0 & \lambda_{53} & 0 \\ 0 & 0 & \lambda_{63} & 0 \\ 0 & 0 & 0 & \lambda_{74} \\ 0 & 0 & 0 & \lambda_{84} \\ 0 & 0 & 0 & \lambda_{94} \end{bmatrix} \begin{bmatrix} \mathrm{KD} \\ \mathrm{BP} \\ \mathrm{OB} \\ \mathrm{LIP} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix}.$$

$$\begin{aligned} \mathrm{KD} = {} & b_1 d_1 + b_2 d_2 + \gamma_1 \mathrm{BP} + \gamma_2 \mathrm{OB} + \gamma_3 \mathrm{LIP} + \gamma_4 (\mathrm{BP} \times \mathrm{OB}) \\ & + \gamma_5 (\mathrm{BP} \times \mathrm{LIP}) + \gamma_6 (\mathrm{OB} \times \mathrm{LIP}) + \gamma_7 (d_1 \times \mathrm{BP}) \\ & + \gamma_8 (d_1 \times \mathrm{OB}) + \gamma_9 (d_2 \times \mathrm{OB}) + \gamma_{10} (d_2 \times \mathrm{LIP}) + \delta. \end{aligned}$$

To further assess the relationships among LIP and the other covariates and/or latent variables, the following submodel is used.

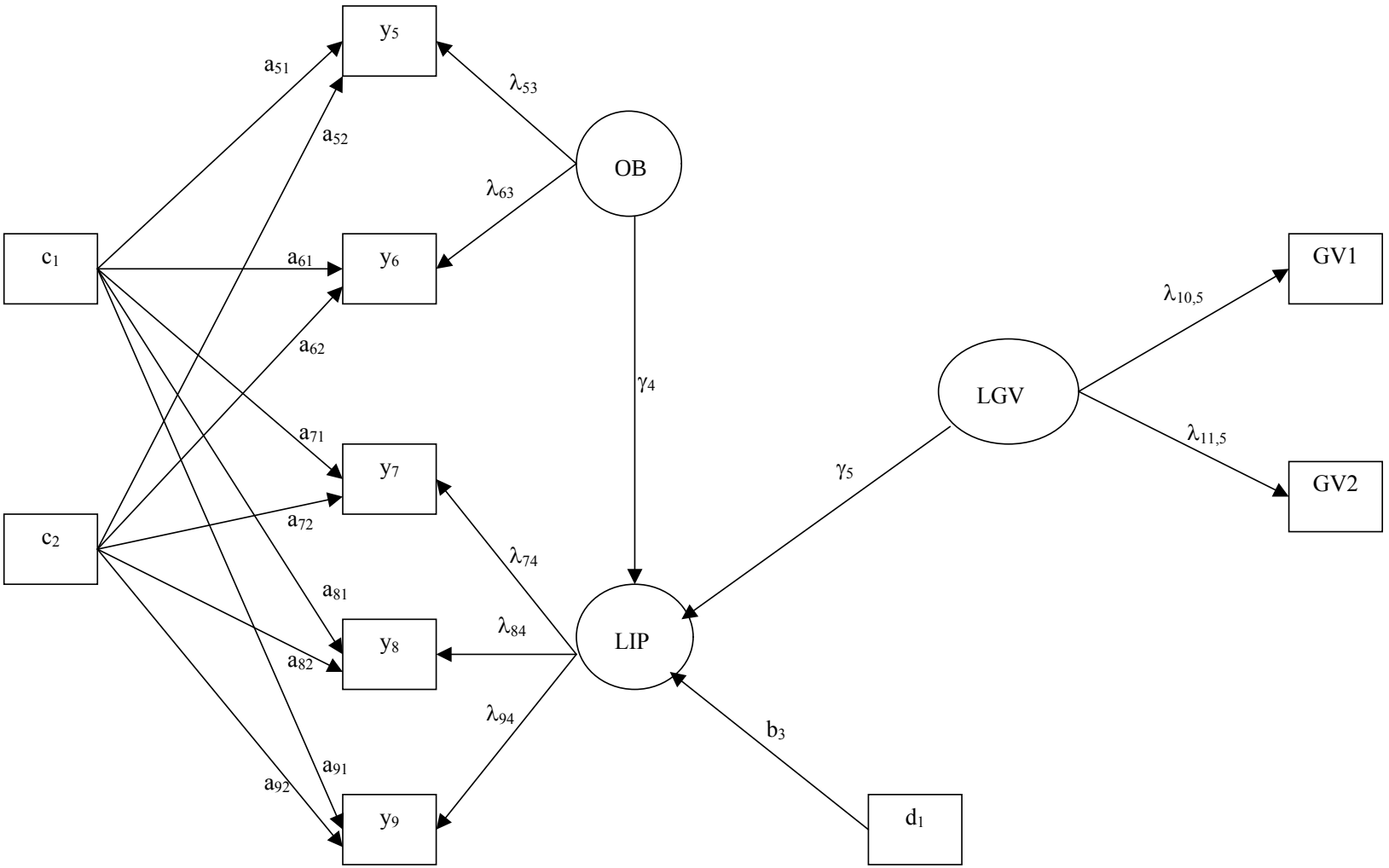
The measurement equation is defined by

$$\begin{bmatrix} y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ \mathrm{GV}_1 \\ \mathrm{GV}_2 \end{bmatrix} = \begin{bmatrix} a_{51} & a_{52} \\ a_{61} & a_{62} \\ a_{71} & a_{72} \\ a_{81} & a_{82} \\ a_{91} & a_{92} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \lambda_{53} & 0 & 0 \\ \lambda_{63} & 0 & 0 \\ 0 & \lambda_{74} & 0 \\ 0 & \lambda_{84} & 0 \\ 0 & \lambda_{94} & 0 \\ 0 & 0 & \lambda_{10,5} \\ 0 & 0 & \lambda_{11,5} \end{bmatrix} \begin{bmatrix} \mathrm{OB} \\ \mathrm{LIP} \\ \mathrm{LGV} \end{bmatrix} + \begin{bmatrix} \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \end{bmatrix}. \quad (29)$$

The structural equation is defined by

$$\mathrm{LIP} = b_3 d_1 + \gamma_4 \mathrm{OB} + \gamma_5 \mathrm{LGV} + \delta. \quad (30)$$

A path diagram representing the submodel defined by (29) and (30) is given below. Here, LIP is treated as an outcome latent variable.



Based on the estimates of γ_3 in the first submodel, and γ_4 in the second submodel, we can obtain an idea about the indirect effect of OB on KD via $\hat{\gamma}_3 \times \hat{\gamma}_4$. However, as $\hat{\gamma}_3$ and $\hat{\gamma}_4$ are not simultaneously estimable through a single model, the estimate of this indirect effect is not optimal and should be interpreted with care.

Note that there are two sets of estimates for the common parameters in these two submodels. However, the differences between them are very small; and in practice, they would not result in different interpretations of the results. Hence, this issue is not important. The trade-off of these disadvantages is the possibility of assessing various nonlinear effects in relation to LIP through one of the submodels.

In this chapter, we discuss the basic SEMs, including

- linear SEMs,
- SEMs with fixed covariates,
- nonlinear SEMs.

Various generalizations will be discussed in subsequent chapters.