Estimating choice models with latent variables with Biogeme

Michel Bierlaire   Moshe Ben-Akiva   Joan Walker

May 13, 2025

Report TRANSP-OR xxxxxx
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

Series on Biogeme

The package Biogeme (`biogeme.epfl.ch`) is designed to estimate the parameters of various models using maximum likelihood estimation. It is particularly designed for discrete choice models. In this document, we present how to estimate choice models involving latent variables.

We assume that the reader is already familiar with discrete choice models, and has successfully installed Biogeme. This document has been written using Biogeme 3.3.

1 Models and notations

The literature on discrete choice models with latent variables is vast (Walker, 2001, Ashok et al., 2002, Greene and Hensher, 2003, Ben-Akiva et al., 2002, to cite just a few). We start this document by a short introduction to the models and the notations.

A *latent variable* is a variable that cannot be directly observed. It is typically modeled using a **structural equation**, which expresses the latent variable as a function of observed (explanatory) variables and an error term. A general form of such a structural equation is:

$$x^*_{nk} = x^*(x_n; \psi_k) + \omega_{nk}, \tag{1}$$

where n indexes individuals, x^*_{nk} is the kth latent variable of interest, x_n is a vector of observed explanatory variables, ψ_k is a vector of parameters to be estimated, and ω_{nk} is a stochastic error term, normally distributed $N(0, \Sigma_{\omega k})$, where $\Sigma_{\omega k}$ is the variance-covariance matrix.

A common specification assumes a linear functional form i.i.d. error terms:

$$x^*_{nk} = \psi_{0k} + \sum_s \psi_{sk} x_{ns} + \sigma_{\omega k}\omega_{nk}, \tag{2}$$

where $\omega_{nk} \sim N(0, 1)$, ψ_{0k} is an intercept term, and $\sigma_{\omega k}$ is a scaling parameter for the error term. The vector $\sigma_\omega = (\sigma_{\omega 1}, \ldots, \sigma_{\omega K})^\top$ corresponds to the diagonal of the covariance matrix $\Sigma_{\omega n}$, with all off-diagonal elements set to zero, implying uncorrelated errors across alternatives.

In discrete choice models, for example, the utility U_{in} that individual n associates with alternative i is a typical example of a latent variable.

Information about latent variables is obtained indirectly through *measurements*, which are observable manifestations of the underlying latent constructs. For example, in discrete choice models, utility is not directly observed but is inferred from the choices individuals make. The relationship between a latent variable and its associated measurements is described by

measurement equations. The specific form of these equations depends on the nature of the observed measurements (e.g., continuous, or ordinal).

1.1 Measurement equations: the continuous case

Since latent variables cannot be directly observed, analysts rely on indirect measurements to infer their values. A common approach involves asking respondents to rate the perceived magnitude of the latent construct on an arbitrary scale. For example: *"How would you rate the level of pain that you are experiencing, from 0 (no pain) to 10 (worst pain imaginable)?"*

Each such rating is referred to as an *indicator*, indexed by $\ell = 1, \ldots, L_n$, and is modeled using a **measurement equation**. This equation relates the observed indicator to the latent variables and other explanatory variables:

$$I_{n\ell} = I_\ell(x_n, x_n^*; \lambda_\ell) + \upsilon_{n\ell}, \ \forall \ell = 1, \ldots, L_n, \forall n, \tag{3}$$

where $I_{n\ell}$ denotes the response provided by individual n for indicator ℓ, x_n^* is the latent variable of interest (e.g., pain perception), x_n is a vector of observed explanatory variables (such as socio-demographic characteristics), λ_ℓ is a vector of K^λ parameters to be estimated, and $\upsilon_{n\ell}$ is a normally distributed random error term with mean 0 and variance-covariance matrix $\Sigma_{\upsilon\ell}$.

A common specification of the measurement function assumes linearity and normally distributed errors:

$$I_{n\ell} = \sum_k \lambda_{\ell k} x_{nk}^* + \sum_s \lambda_{\ell s} x_{ns} + \sigma_{\upsilon\ell}\upsilon_{n\ell}, \quad \forall \ell, \tag{4}$$

where $\lambda_{\ell k}$ are unknown parameters to be estimated, $\sigma_{\upsilon\ell}$ is an indicator-specific scale parameter, and $\upsilon_{n\ell} \sim N(0, 1)$.

1.2 Measurement equation: the ordinal case

Another type of indicator arises when respondents are asked to evaluate a statement using an ordinal scale. A typical context for this type of measurement is the use of a Likert scale (Likert, 1932), where individuals express their degree of agreement or disagreement with a given statement. For example:

> *"I believe that my own actions have an impact on the planet."*
> Response options: strongly agree (2), agree (1), neutral (0), disagree (−1), strongly disagree (−2).

Another common example is the observed choice itself. In discrete choice models, whether or not an alternative is chosen is represented by a binary variable, which can be interpreted as a special case of an ordinal scale with only two categories.

To model these types of indicators, we represent the observed measurement as an *ordered discrete variable* $I_{n\ell}$, which takes values in a finite, ordered set $\{j_1, j_2, \ldots, j_{M_\ell}\}$. The measurement equation involves two stages:

Step 1: Latent response formulation. We first define a continuous response variable, that happens to be unobserved (latent) in this case:

$$I^*_{n\ell} = I^*_\ell(x_n, x^*_n; \lambda_\ell) + \upsilon_{n\ell}, \tag{5}$$

where $I^*_{n\ell}$ is a continuous latent variable underlying the reported response, x^*_n is the relevant latent variable (e.g., environmental concern), x_n is a vector of observed explanatory variables (e.g., age, income), λ is a vector of parameters to be estimated, and $\upsilon_{n\ell}$ is a random error term.

Step 2: Discretization via thresholds. Since $I^*_{n\ell}$ is not observed, we relate it to the reported discrete measurement $I_{n\ell}$ through a set of threshold parameters:

$$I_{n\ell} = \begin{cases} j_1 & \text{if } I^*_{n\ell} < \tau_1, \\ j_2 & \text{if } \tau_1 \leqslant I^*_{n\ell} < \tau_2, \\ \vdots & \\ j_m & \text{if } \tau_{m-1} \leqslant I^*_{n\ell} < \tau_m, \\ \vdots & \\ j_M & \text{if } \tau_{M_\ell-1} \leqslant I^*_{n\ell}, \end{cases} \tag{6}$$

where $\tau_1, \ldots, \tau_{M_\ell-1}$ are threshold parameters to be estimated, satisfying the ordering constraint:

$$\tau_1 \leqslant \tau_2 \leqslant \cdots \leqslant \tau_{M_\ell-1}. \tag{7}$$

Note that it is customary to use the same set of parameters for all individuals n and all indicators ℓ, which explains the absence of these indices on the parameter τ.

Defining $\tau_0 = -\infty$ and $\tau_{M_\ell} = +\infty$, it simplifies to

$$I_{n\ell} = j_m \text{ if } \tau_{m-1} \leqslant I^*_{n\ell} < \tau_m,\ m = 1, \ldots, M_\ell. \tag{8}$$

1.3 Summary of notations

n	index for individuals	$\mathbb{N}$
N	number of individuals in the sample	$\mathbb{N}$
$\mathcal{C}$	universal choice set	
$\mathcal{C}_n$	choice set of individual n	
$i, j \in \mathcal{C}_n$	indices of alternatives	
J	total number of alternatives in $\mathcal{C}$	$\mathbb{N}_0$
J_n	number of alternatives in $\mathcal{C}_n$	$\mathbb{N}$
L_n	number of indicators available for individal n	$\mathbb{N}$
M_ℓ	number of levels for the discrete Likert scale of indicator ℓ	$\mathbb{N}$
$\theta = (\theta_1, \dots, \theta_{K^\theta})^T$	vector of all unknown parameters	$\mathbb{R}^{K^\theta}$
$\beta = (\beta_1, \dots, \beta_K)^T$	vector of unknown coefficients in the systematic part of the utility	$\mathbb{R}^K$
$\gamma = (\gamma_1, \dots, \gamma_{K^\gamma})^T$	vector of unknown parameters that are not coefficients	$\mathbb{R}^{K^\gamma}$
$\alpha = (\alpha_1, \dots, \alpha_{K^\alpha})^T$	vector of all unknown parameters of structural equations	$\mathbb{R}^{K^\alpha}$
$\psi = (\psi_1, \dots, \psi_{K^\psi})^T$	vector of unknown coefficients in the systematic part of the structural equations	$\mathbb{R}^{K^\alpha}$
$\kappa = (\kappa_1, \dots, \kappa_{K^\kappa})^T$	vector of all unknown parameters of measurement equations	$\mathbb{R}^{K^\alpha}$
$\lambda = (\lambda_1, \dots, \lambda_{K^\lambda})^T$	vector of unknown coefficients in the systematic part of measurement equations	$\mathbb{R}^{K^\lambda}$
$m = (m_1, \dots, m_{K^m})^T$	vector of measurements/indicators	$\mathbb{R}^{K^m}$
S_n	vector of characteristics of individual n or the choice context	
z_{in}	vector of attributes describing alternative i as perceived by individual n	

$x_{in} = (x_{in1}, \ldots, x_{inK^x})^T$	vector of explanatory variables for alternative i and individual n, function of attributes and characteristics, that is $x_{in} = h(z_{in}, S_n)$ (for notational simplification, K^x may be denoted by K or L in the text)	$\mathbb{R}^{K^x}$
$x_n^* = (x_{n1}, \ldots, x_{nK^{x^*}})^T$	vector of latent variables for individual n	$\mathbb{R}^{K^{x^*}}$
ε_n	vector of error terms of the utility functions	$\mathbb{R}$
ω_n	vector of error terms of structural equations	
υ_n	vector of error terms of measurement equations	
$\mathcal{L}^*(\theta)$	likelihood function	$\mathbb{R}^{K^\theta} \to \mathbb{R}$
$\mathcal{L}(\theta)$	log-likelihood function, $\mathcal{L}(\theta) = \ln \mathcal{L}^*(\theta)$	$\mathbb{R}^{K^\theta} \to \mathbb{R}$,
$f_\varepsilon(\cdot;\theta)$	probability density function (pdf) of continuous random variable ε, parameterized by the vector θ	$\mathbb{R}^{K^\varepsilon} \to \mathbb{R}^+$
$F_\varepsilon(\cdot;\theta)$	cumulative distribution function (CDF) of random variable ε, parameterized by θ	$\mathbb{R}^{K^\varepsilon} \to [0,1]$
$\phi(\cdot)$	probability density function of the univariate standardized normal distribution	$\mathbb{R} \to \mathbb{R}^+$
$\Phi(\cdot)$	cumulative distribution function of the univariate standardized normal distribution	$\mathbb{R} \to [0,1]$
R	number of draws in simulation context	$\mathbb{N}$
r	index of the draws in simulation context	$\mathbb{N}$
Σ_ξ	variance-covariance matrix of the normal random vector ξ	$\mathbb{R}^{K^\xi \times K^\xi}$

2 The MIMIC model

The Multiple Indicators Multiple Causes (MIMIC) model is a structural equation modeling framework designed to analyze relationships involving latent variables. In a MIMIC model, the latent variable is simultaneously influenced by a set of observed explanatory variables (the "multiple causes") and reflected in several observed indicators (the "multiple indicators"). This dual structure enables the analyst to capture both the determinants and the manifestations of latent constructs, such as attitudes, preferences, or psychological traits. A seminal introduction to the MIMIC model is provided by Jöreskog and Goldberger (1975), who formalized its use within the broader class of structural equation models.

The model involves the structural equations (1) and the measurement equations (3) and (6). Conditional on the latent variables, the contribution of each indicator ℓ for each observation n to the likelihood function is defined as follows.

$$\begin{aligned}\Pr(I_{n\ell} = j_m | x_n^*, x_n; \lambda_\ell, \Sigma_{\upsilon\ell}) &= \Pr(\tau_{m-1} < I_{n\ell}^* \leqslant \tau_m) \\ &= \Pr(I_{n\ell}^* \leqslant \tau_m) - \Pr(I_{n\ell}^* \leqslant \tau_{m-1}),\end{aligned} \tag{9}$$

where j_m category j_m observed is then given by:

We can assume a linear specification of the latent response function, as in Equation (4):

$$I_{n\ell}^* = \sum_k \lambda_{\ell k} x_{nk}^* + \sum_s \lambda_{\ell s} x_{ns} + \sigma_{\upsilon\ell}\upsilon_{n\ell}, \quad \forall j, \tag{10}$$

where the error term $\upsilon_{n\ell} \sim \mathsf{N}(0, 1)$. Using (2), we obtain

$$\begin{aligned}I_{n\ell}^* &= \sum_k \lambda_{\ell k} x_{nk}^* + \sum_s \lambda_{\ell k} x_{ns} + \sigma_{\upsilon\ell}\upsilon_{n\ell} \\ &= \sum_k \lambda_{\ell k}\psi_{0k} + \sum_k \lambda_{\ell k} \sum_s \psi_{sk} x_{ns} + \sum_k \lambda_k \sigma_{\omega k}\omega_{nk} + \sum_s \lambda_{\ell s} x_{ns} + \sigma_{\upsilon\ell}\upsilon_{n\ell}\end{aligned}$$

Clearly, the parameters of the structural and the measurement equations are confounded. The equation can be written

$$I_{n\ell}^* = \lambda'_{\ell 0} + \sum_s \lambda'_{\ell s} x_{ns} + \sigma'_{\upsilon\ell}\upsilon'_{n\ell}, \tag{11}$$

where $\lambda'_{\ell 0} = \sum_k \lambda_{\ell k}\psi_{0k}$ is the intercept, $\lambda'_{\ell s} = \sum_k \lambda_{\ell k}\psi_{sk} + \lambda_{\ell s}$ are the coefficients of the socio-economic characteristics, $\sigma'_\ell = \sqrt{(\sum_k \lambda_k \sigma_{\omega k})^2 + \sigma_{\upsilon\ell}^2}$ is the scale parameters, and $\upsilon'_{n\ell} \sim \mathsf{N}(0, 1)$.

The probability of observing category j_i for individual n and indicator ℓ becomes:

$$
\begin{aligned}
\Pr(I_{n\ell} = j_m | x^*_n, x_n; \lambda_\ell, \Sigma_{\upsilon\ell}) &= \Pr(I_{n\ell} = j_m | x_n; \lambda'_\ell, \sigma'_{\upsilon\ell}) \\
&= \Pr(I^*_{n\ell} \leqslant \tau_m) - \Pr(I^*_{n\ell} \leqslant \tau_{m-1}) \\
&= \Pr\left(\upsilon'_{n\ell} \leqslant \frac{\tau_m - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \\
&\quad - \Pr\left(\upsilon'_{n\ell} \leqslant \frac{\tau_{m-1} - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \\
&= \Phi\left(\frac{\tau_m - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \\
&\quad - \Phi\left(\frac{\tau_{m-1} - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right),
\end{aligned}
$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the standard normal distribution. x This specification is known as the *ordered probit model* and is widely used for modeling ordinal responses that depend on latent constructs.

The likelihood function is therefore:

$$
\mathcal{L}^*(\lambda', \sigma', \tau) = \prod_n \prod_\ell \prod_m \left(\begin{array}{c} \Phi\left(\frac{\tau_m - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \\ - \Phi\left(\frac{\tau_{m-1} - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \end{array} \right),
$$

where $\tau_0 = -\infty$ and $\tau_{M_\ell} = +\infty$. Therefore, the log-likelihood function $\log \mathcal{L}^*(\lambda', \sigma')$ is

$$
\mathcal{L}(\lambda', \sigma', \tau) = \sum_n \sum_\ell \sum_m \log \left(\begin{array}{c} \Phi\left(\frac{\tau_m - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \\ - \Phi\left(\frac{\tau_{m-1} - \lambda'_{\ell 0} - \sum_s \lambda'_{\ell s} x_{ns}}{\sigma'_{\upsilon\ell}}\right) \end{array} \right).
$$

Once the parameters λ' and σ' have been estimated, there are infinitely many combinations of the original parameters that verify the equations

$$
\begin{aligned}
\lambda'_{\ell 0} &= \sum_k \lambda_{\ell k} \psi_{0k}, \\
\lambda'_{\ell s} &= \sum_k \lambda_{\ell k} \psi_{sk} + \lambda_{\ell s}, \\
\sigma'_\ell &= \sqrt{(\sum_k \lambda_k \sigma_{\omega k})^2 + \sigma^2_{\upsilon\ell}}.
\end{aligned}
$$

It is therefore necessary to normalize the specification. A common practice is to associate each latent variable k with an indicator ℓ, and set $\lambda_{0k} = 0$, $\lambda_{\ell k} = 1$ and $\sigma_{\upsilon\ell} = 0$.

Consider an example where there is only one latent variable x^*_n, two explanatory variables and three indicators. The structural equation (2) is written

$$x^*_n = \psi_0 + \psi_1 x_{n1} + \psi_2 x_{n2} + \sigma_\omega \omega_{nk}.$$

The measurement equations (10) are written

$$\begin{aligned}
I^*_{n1} &= \lambda_{10} x^*_n + \lambda_{11} x_{n1} + \lambda_{12} x_{n2} + \sigma_{\upsilon 1} \upsilon_{n1},\\
I^*_{n2} &= \lambda_{20} x^*_n + \lambda_{21} x_{n1} + \lambda_{22} x_{n2} + \sigma_{\upsilon 2} \upsilon_{n2},\\
I^*_{n3} &= \lambda_{30} x^*_n + \lambda_{31} x_{n1} + \lambda_{32} x_{n2} + \sigma_{\upsilon 3} \upsilon_{n3}.
\end{aligned}$$

Substituting x^*_n, we obtain

$$\begin{aligned}
I^*_{n1} &= \lambda_{10}\psi_0 + (\lambda_{10}\psi_1 + \lambda_{11}) x_{n1} + (\lambda_{10}\psi_2 + \lambda_{12}) x_{n2} + \lambda_{10}\sigma_\omega \omega_n + \sigma_{\upsilon 1}\upsilon_{n1},\\
&= \lambda'_{10} + \lambda'_{11} x_{n1} + \lambda'_{12} x_{n2} + \sigma'_{\upsilon 1} \upsilon'_{n1},\\
I^*_{n2} &= \lambda_{20}\psi_0 + (\lambda_{20}\psi_1 + \lambda_{21}) x_{n1} + (\lambda_{20}\psi_2 + \lambda_{22}) x_{n2} + \lambda_{20}\sigma_\omega \omega_n + \sigma_{\upsilon 1}\upsilon_{n1},\\
&= \lambda'_{20} + \lambda'_{21} x_{n1} + \lambda'_{22} x_{n2} + \sigma'_{\upsilon 2} \upsilon'_{n2},\\
I^*_{n3} &= \lambda_{30}\psi_0 + (\lambda_{30}\psi_1 + \lambda_{31}) x_{n1} + (\lambda_{30}\psi_2 + \lambda_{32}) x_{n2} + \lambda_{30}\sigma_\omega \omega_n + \sigma_{\upsilon 1}\upsilon_{n1},\\
&= \lambda'_{30} + \lambda'_{31} x_{n1} + \lambda'_{32} x_{n2} + \sigma'_{\upsilon 3} \upsilon'_{n3},
\end{aligned}$$

where

$$\begin{aligned}
\lambda'_{\ell 0} &= \lambda_{\ell 0}\psi_0,\\
\lambda'_{\ell 1} &= \lambda_{\ell 0}\psi_1 + \lambda_{\ell 1},\\
\lambda'_{\ell 2} &= \lambda_{\ell 0}\psi_2 + \lambda_{\ell 2},\\
\sigma'_{\upsilon\ell} &= \sqrt{(\lambda_{\ell 0}\sigma_\omega)^2 + (\sigma_{\upsilon\ell})^2}.
\end{aligned}$$

Now, if we associate the latent variable with the first indicator, it means that we normalize $\lambda_{10} = 1$, $\lambda_{11} = \lambda_{12} = 0$ and $\sigma_{\upsilon 1} = 0$, so that the first measurement equation is now written

$$I^*_{n1} = x^*_n.$$

Once the parameters λ' and σ' have been estimated, the original parameters can be recovered by solving the following system of equations:

$$\begin{aligned}
\lambda'_{10} &= \psi_0,\\
\lambda'_{20} &= \lambda_{20}\psi_0,\\
\lambda'_{30} &= \lambda_{30}\psi_0,\\
\lambda'_{11} &= \psi_1,\\
\lambda'_{21} &= \lambda_{20}\psi_1 + \lambda_{21},\\
\lambda'_{31} &= \lambda_{30}\psi_1 + \lambda_{31},\\
\lambda'_{12} &= \psi_2,\\
\lambda'_{22} &= \lambda_{20}\psi_2 + \lambda_{22},\\
\lambda'_{32} &= \lambda_{30}\psi_2 + \lambda_{32},\\
\sigma'_{\upsilon 1} &= \sigma_\omega,\\
\sigma'_{\upsilon 2} &= \sqrt{(\lambda_{20}\sigma_\omega)^2 + (\sigma_{\upsilon 2})^2},\\
\sigma'_{\upsilon 3} &= \sqrt{(\lambda_{30}\sigma_\omega)^2 + (\sigma_{\upsilon 3})^2}.
\end{aligned}$$

Consider now a more complex example where there are two latent variables x^*_{n1} and x^*_{n2}, one explanatory variable x_n and four indicators. The structural equations (2) are written

$$\begin{aligned}
x^*_{n1} &= \psi_{01} + \psi_{11}x_n + \sigma_{\omega 1}\omega_{n1},\\
x^*_{n2} &= \psi_{02} + \psi_{12}x_n + \sigma_{\omega 2}\omega_{n2}.
\end{aligned}$$

The measurement equations (10) are written

$$\begin{aligned}
I^*_{n1} &= \lambda_{10}x^*_{n1} + \lambda_{11}x^*_{n2} + \lambda_{12}x_n + \sigma_{\upsilon 1}\upsilon_{n1},\\
I^*_{n2} &= \lambda_{20}x^*_{n1} + \lambda_{21}x^*_{n2} + \lambda_{22}x_n + \sigma_{\upsilon 2}\upsilon_{n2},\\
I^*_{n3} &= \lambda_{30}x^*_{n1} + \lambda_{31}x^*_{n2} + \lambda_{32}x_n + \sigma_{\upsilon 3}\upsilon_{n3}.\\
I^*_{n4} &= \lambda_{40}x^*_{n1} + \lambda_{41}x^*_{n2} + \lambda_{42}x_n + \sigma_{\upsilon 4}\upsilon_{n4}.
\end{aligned}$$

Substituting x^*_n, we obtain

$$\begin{aligned}
I^*_{n1} =&(\lambda_{10}\psi_{01} + \lambda_{11}\psi_{02}) + (\lambda_{10}\psi_{11} + \lambda_{11}\psi_{12} + \lambda_{12})x_n +\\
&\lambda_{10}\sigma_{\omega 1}\omega_{n1} + \lambda_{11}\sigma_{\omega 2}\omega_{n2} + \sigma_{\upsilon 1}\upsilon_{n1},\\
I^*_{n2} =&(\lambda_{20}\psi_{01} + \lambda_{21}\psi_{02}) + (\lambda_{20}\psi_{11} + \lambda_{21}\psi_{12} + \lambda_{22})x_n +\\
&\lambda_{20}\sigma_{\omega 1}\omega_{n1} + \lambda_{21}\sigma_{\omega 2}\omega_{n2} + \sigma_{\upsilon 1}\upsilon_{n1},\\
I^*_{n3} =&(\lambda_{30}\psi_{01} + \lambda_{31}\psi_{02}) + (\lambda_{30}\psi_{11} + \lambda_{31}\psi_{12} + \lambda_{32})x_n +\\
&\lambda_{30}\sigma_{\omega 1}\omega_{n1} + \lambda_{31}\sigma_{\omega 2}\omega_{n2} + \sigma_{\upsilon 1}\upsilon_{n1}.\\
I^*_{n4} =&(\lambda_{40}\psi_{01} + \lambda_{41}\psi_{02}) + (\lambda_{40}\psi_{11} + \lambda_{41}\psi_{12} + \lambda_{42})x_n +\\
&\lambda_{40}\sigma_{\omega 1}\omega_{n1} + \lambda_{41}\sigma_{\omega 2}\omega_{n2} + \sigma_{\upsilon 1}\upsilon_{n1}.
\end{aligned}$$

Defining

$$\begin{aligned}
\lambda'_{10} &= \lambda_{10}\psi_{01} + \lambda_{11}\psi_{02},\\
\lambda'_{12} &= (\lambda_{10}\psi_{11} + \lambda_{11}\psi_{12} + \lambda_{12})\\
\sigma'_{\upsilon 1} &= \sqrt{(\lambda_{10}\sigma_{\omega 1})^2 + (\lambda_{11}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 1}},\\
\lambda'_{20} &= \lambda_{20}\psi_{01} + \lambda_{21}\psi_{02},\\
\lambda'_{22} &= (\lambda_{20}\psi_{11} + \lambda_{21}\psi_{12} + \lambda_{22})\\
\sigma'_{\upsilon 2} &= \sqrt{(\lambda_{20}\sigma_{\omega 1})^2 + (\lambda_{21}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 2}},\\
\lambda'_{30} &= \lambda_{30}\psi_{01} + \lambda_{31}\psi_{02},\\
\lambda'_{32} &= (\lambda_{30}\psi_{11} + \lambda_{31}\psi_{12} + \lambda_{32})\\
\sigma'_{\upsilon 3} &= \sqrt{(\lambda_{30}\sigma_{\omega 1})^2 + (\lambda_{31}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 3}},\\
\lambda'_{40} &= \lambda_{40}\psi_{01} + \lambda_{41}\psi_{02},\\
\lambda'_{42} &= (\lambda_{40}\psi_{11} + \lambda_{41}\psi_{12} + \lambda_{42})\\
\sigma'_{\upsilon 4} &= \sqrt{(\lambda_{40}\sigma_{\omega 1})^2 + (\lambda_{41}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 4}},
\end{aligned}$$

we obtain

$$\begin{aligned}
I^*_{n1} =& \lambda'_{10} + \lambda'_{12}x_n + \sigma'_{\upsilon 1}\upsilon'_{n1},\\
I^*_{n2} =& \lambda'_{20} + \lambda'_{22}x_n + \sigma'_{\upsilon 2}\upsilon'_{n2},\\
I^*_{n3} =& \lambda'_{30} + \lambda'_{32}x_n + \sigma'_{\upsilon 3}\upsilon'_{n3}, I^*_{n4} = \quad \lambda'_{40} + \lambda'_{42}x_n + \sigma'_{\upsilon 4}\upsilon'_{n4}.
\end{aligned}$$

Now, if we associate the first latent variable x^*_{n1} with the first indicator I^*_{n1}, and the second latent variable x^*_{n2} with the second indicator I^*_{n2}, it is equivalent to normalize the parameters $\lambda_{10} = 1$, $\lambda_{11} = \lambda_{12} = \sigma_{\upsilon 1} = 0$ for the first indicator, and the parameters $\lambda_{21} = 1$ and $\lambda_{20} = \lambda_{22} = \sigma_{\upsilon 2} = 0$ for the second one. The measurement equations then become

$$\begin{aligned}
I^*_{n1} &= x^*_{n1},\\
I^*_{n2} &= x^*_{n2},\\
I^*_{n3} &= \lambda_{30}x^*_{n1} + \lambda_{31}x^*_{n2} + \lambda_{32}x_n + \sigma_{\upsilon 3}\upsilon_{n3},\\
I^*_{n4} &= \lambda_{40}x^*_{n1} + \lambda_{41}x^*_{n2} + \lambda_{42}x_n + \sigma_{\upsilon 4}\upsilon_{n4}.
\end{aligned}$$

Once the parameters λ' and σ' have been estimated, the original parameters can be recovered by solving the following system of equations:

$$\lambda'_{10} = \psi_{01},$$
$$\lambda'_{12} = \psi_{11},$$
$$\sigma'_{\upsilon 1} = \sigma_{\omega 1},$$
$$\lambda'_{20} = \psi_{02},$$
$$\lambda'_{22} = \psi_{12},$$
$$\sigma'_{\upsilon 2} = \sigma_{\omega 2},$$
$$\lambda'_{30} = \lambda_{30}\psi_{01} + \lambda_{31}\psi_{02},$$
$$\lambda'_{32} = (\lambda_{30}\psi_{11} + \lambda_{31}\psi_{12} + \lambda_{32})$$
$$\sigma'_{\upsilon 3} = \sqrt{(\lambda_{30}\sigma_{\omega 1})^2 + (\lambda_{31}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 3}},$$
$$\lambda'_{40} = \lambda_{40}\psi_{01} + \lambda_{41}\psi_{02},$$
$$\lambda'_{42} = (\lambda_{40}\psi_{11} + \lambda_{41}\psi_{12} + \lambda_{42})$$
$$\sigma'_{\upsilon 4} = \sqrt{(\lambda_{40}\sigma_{\omega 1})^2 + (\lambda_{41}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 4}},$$

that is

$$\lambda'_{10} = \psi_{01},$$
$$\lambda'_{12} = \psi_{11},$$
$$\sigma'_{\upsilon 1} = \sigma_{\omega 1},$$
$$\lambda'_{20} = \psi_{02},$$
$$\lambda'_{22} = \psi_{12},$$
$$\sigma'_{\upsilon 2} = \sigma_{\omega 2},$$
$$\lambda'_{30} = \lambda_{30}\psi_{01} + \lambda_{31}\psi_{02},$$
$$\lambda'_{32} = (\lambda_{30}\psi_{11} + \lambda_{31}\psi_{12} + \lambda_{32}$$
$$\sigma'_{\upsilon 3} = \sqrt{(\lambda_{30}\sigma_{\omega 1})^2 + (\lambda_{31}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 3}},$$
$$\lambda'_{40} = \lambda_{40}\psi_{01} + \lambda_{41}\psi_{02},$$
$$\lambda'_{42} = (\lambda_{40}\psi_{11} + \lambda_{41}\psi_{12} + \lambda_{42}$$
$$\sigma'_{\upsilon 4} = \sqrt{(\lambda_{40}\sigma_{\omega 1})^2 + (\lambda_{41}\sigma_{\omega 2})^2 + \sigma^2_{\upsilon 4}},$$

MIMIC Model with Cross-Loading

We consider a MIMIC model with K^* latent variables, L indicators, and K explanatory variables. The model is defined by:

Structural Equations (Latent Variables)

$$z_{nk^*} = \psi_{0k^*} + \sum_{k=1}^{K} \psi_{kk^*} x_{nk} + \sigma_{\omega,k^*} \cdot \omega_{nk^*}, \quad \omega_{nk^*} \sim \mathcal{N}(0,1)$$

Measurement Equations (Indicators)

$$y_{n\ell} = \lambda_{\ell 0} + \sum_{k^*=1}^{K^*} \lambda_{\ell k^*} z_{nk^*} + \sigma_{\epsilon,\ell} \cdot \epsilon_{n\ell}, \quad \epsilon_{n\ell} \sim \mathcal{N}(0,1)$$

When an indicator is associated with multiple latent variables (cross-loading), identifiability requires proper normalization.

Normalization Strategy

To ensure identifiability:

- Fix the location and scale of each latent variable.
- Either:
  - Fix one intercept and one loading per latent variable, e.g., $\lambda_{\ell_0 k^*} = 1$ and $\lambda_{\ell_0 0} = 0$.
  - Or fix the variance of each latent variable: $\mathrm{Var}(z_{k^*}) = 1$ and fix at least one loading per latent.
- Impose enough constraints to avoid rotational indeterminacy.

Toy Numerical Example

Let us consider:

- $K^* = 2$ latent variables: z_1, z_2
- $K = 2$ observed causes: x_1, x_2
- $L = 3$ indicators: y_1, y_2, y_3

Structural Equations

$$z_1 = 0.5 + 1.0x_1 + 0.2x_2$$
$$z_2 = -0.3 + 0.4x_1 + 0.9x_2$$

Measurement Equations (with Cross-Loading)

$$
\begin{aligned}
y_1 &= z_1 + \epsilon_1 \\
y_2 &= z_2 + \epsilon_2 \\
y_3 &= 0.5z_1 + 0.7z_2 + \epsilon_3
\end{aligned}
$$

Normalization Used

- $\lambda_{11} = 1$, $\lambda_{10} = 0$ (for z_1)
- $\lambda_{22} = 1$, $\lambda_{20} = 0$ (for z_2)

Prediction Example

Let $x_1 = 2$, $x_2 = 1$. Then:

$$
\begin{aligned}
z_1 &= 0.5 + 1.0 \cdot 2 + 0.2 \cdot 1 = 2.7 \\
z_2 &= -0.3 + 0.4 \cdot 2 + 0.9 \cdot 1 = 1.4
\end{aligned}
$$

Predicted indicators:

$$
\begin{aligned}
\hat{y}_1 &= z_1 = 2.7 \\
\hat{y}_2 &= z_2 = 1.4 \\
\hat{y}_3 &= 0.5 \cdot z_1 + 0.7 \cdot z_2 = 0.5 \cdot 2.7 + 0.7 \cdot 1.4 = 1.35 + 0.98 = 2.33
\end{aligned}
$$

Conclusion

The normalization ensures identifiability even when indicators depend on multiple latent variables. Once normalized, predictions can be made using the structural part of the model.

3 Rotational Indeterminacy

In models with multiple latent variables—such as MIMIC or factor models—a key issue is **rotational indeterminacy**. This means that the estimated latent variables and their loadings are not uniquely determined. Any orthogonal rotation of the latent space leads to an equivalent model in terms of fit and likelihood.

3.1 2. Measurement Model Setup

Let $\mathbf{z}_n \in \mathbb{R}^{K^*}$ denote a vector of K^* latent variables for observation n, and let $\mathbf{y}_n \in \mathbb{R}^L$ be a vector of L observed indicators. The measurement model is:

$$\mathbf{y}_n = \Lambda \mathbf{z}_n + \boldsymbol{\epsilon}_n, \quad \boldsymbol{\epsilon}_n \sim \mathcal{N}(0, \Sigma_\epsilon)$$

where $\Lambda \in \mathbb{R}^{L \times K^*}$ is the loading matrix.

3.2 3. Rotational Indeterminacy

Suppose $R \in \mathbb{R}^{K^* \times K^*}$ is an **orthogonal matrix**, meaning $R^\top R = I$. Define new latent variables:

$$\tilde{\mathbf{z}}_n = R\mathbf{z}_n$$

Then the model becomes:

$$\mathbf{y}_n = \Lambda R^{-1} \tilde{\mathbf{z}}_n + \boldsymbol{\epsilon}_n$$

Let $\tilde{\Lambda} = \Lambda R^{-1}$. The model:

$$\mathbf{y}_n = \tilde{\Lambda} \tilde{\mathbf{z}}_n + \boldsymbol{\epsilon}_n$$

is **observationally equivalent**. The likelihood, residuals, and predictions are the same.

3.3 4. Geometric Interpretation

Rotating the latent variables corresponds to rotating the coordinate system in latent space. Since the model depends only on the linear combination $\Lambda \mathbf{z}_n$, any rotation of $\mathbf{z}_n$ can be absorbed into Λ without changing the result.

3.4 5. Why It's a Problem

- **Non-unique parameters:** The loadings and latent scores are not uniquely defined.
- **Interpretability:** The meaning of each latent variable becomes arbitrary.
- **Estimation instability:** The optimizer may converge to different solutions depending on initialization or constraints.

3.5 6. How to Resolve It

To identify the model, we impose constraints:

- **Fix some loadings**, e.g., $\lambda_{\ell k^*} = 1$ for one ℓ per k^*.
- **Fix the variance** of each latent variable: $\mathrm{Var}(z_{k^*}) = 1$.
- **Force a simple structure**, e.g., many loadings are zero.

3.6 7. Example: Rotation Invariance

Let us consider two latent variables z_1 and z_2, and two indicators:

$$y_1 = z_1 + z_2$$
$$y_2 = z_1 - z_2$$

Define the orthogonal rotation matrix:

$$R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Then:

$$\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix} = R \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} z_1 + z_2 \\ -z_1 + z_2 \end{bmatrix}$$

Now define new loadings:

$$\tilde{\Lambda} = \Lambda R^{-1}$$

This gives a new model with rotated latent variables and adjusted loadings, but:

$$y_1, y_2 \text{ are unchanged}$$

3.7 8. Conclusion

Rotational indeterminacy implies that without appropriate normalization or constraints, latent variable models are not uniquely identifiable. Understanding and resolving this ambiguity is essential for interpretability and stability.

References

Ashok, K., Dillon, W. R. and Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables, *Journal of Marketing Research* **39**(1): 31–46.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models, *Perpetual motion: Travel behaviour research opportunities and application challenges* pp. 431–470.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit, *Transportation Research Part B: Methodological* **37**(8): 681–698.

Jöreskog, K. G. and 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* **70**(351a): 631–639. DOI: 10.1080/01621459.1975.10482485.

Likert, R. (1932). A technique for the measurement of attitudes, *Archives of psychology* **140**: 1–55.

Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*, PhD thesis, Massachusetts Institute of Technology.