Latent Variable Approach

Jaya Krishnakumar, University of Geneva

Main idea:

- The theoretical concept is not directly observable; it is latent (hidden)
- The observed indicators /outcomes or responses are partial/imperfect measures of the underlying theoretical concept
- How to make inference about the latent variable using the observed indicators?

Example

- Latent variable: the freedom of choice that an individual has in each dimension of human wellbeing (the capability set)
- Observed indicators: the achievement indicators
- Other relevant information: personal characteristics of the individual and some features of the 'world' in which s/he lives

Main Models

- Factor Analysis (FA)
- Multiple Indicators Multiple Causes (MIMIC)
- Structural Equation Models (SEM)
- Extensions with covariates (exogenous variables) and qualitative outcomes

Factor Analysis

- In this model, the observed values are postulated to be (linear) functions of a certain number (fewer) of unobserved latent variables (called factors). These latent factors represent our capabilities.
- However this model does *not* explain the latent variables (or the capabilities) themselves.

Factor Analysis (contd.)

- The Model

$$y = \Lambda f + \varepsilon$$

$$V(f) = \Phi \quad and \quad V(\varepsilon) = \Psi$$

$$\Sigma = \Lambda \Phi \Lambda' + \Psi.$$

Factor Analysis (contd.)

- Maximum likelihood procedure is applied to the model to estimate Λ and Ψ given Σ. Given Λ , Ψ one can derive minimum variance estimators or predictors of *f* as follows:

$$\hat{f} = (I + \Gamma)^{-1} \Lambda' \Psi^{-1} y$$

- Or as below for an unbiased version

$$\hat{f}^* = \Gamma^{-1} \Lambda' \Psi^{-1} y = (\Lambda' \Psi^{-1} \Lambda)^{-1} \Lambda' \Psi^{-1} y$$

Factor Analysis (contd.)

The special case $\Psi = I$

$$\tilde{f} = (I + \Lambda'\Lambda)^{-1}\Lambda'$$

and the ‘unbiased’ version

$$\tilde{f}^* = (\Lambda'\Lambda)^{-1}\Lambda' y$$

Principal Components

- What about principal components (PC)? PC is *not* a latent variable model, but a data reduction technique. This method seeks linear combinations of the observed indicators in such a way as to reproduce the original variance as closely as possible.
- It is widely used in empirical applications as an 'aggregating' technique
- Under certain conditions PC's can be shown to be equivalent to the factor scores obtained in FA

Link between PC and FA

- The estimators of the latent variables obtained above for $\Psi = I$ can be shown to be proportional to the (first *m*) principal components say p^*.
- This identity between the 'unbiased' versions of PC's and factor scores provides the theoretical justification for the possible interpretation of principal components as latent variable estimators under special conditions.

MIMIC Models

- Multiple Indicators Multiple Causes
- This model goes a step further in the explanation.
- Here the observed variables are taken to be manifestations of a latent concept (as in the FA model) but the latent factor(s) are in turn “caused” by exogenous elements (the individual characteristics and the ‘world’ we mentioned earlier).

MIMIC Models (contd.)

According to this model, the observed variables result from the latent factors and the latent factors themselves are caused by other exogenous variables denoted here as . Thus we have a ‘measurement equation' and a ‘causal' relationship:

$$y = \lambda f + \varepsilon$$

$$f = \beta' x + \varepsilon$$

MIMIC Models (contd.)

- The estimator of *f* is given by

$$\hat{f} = (1 - \lambda' \Omega^{-1} \lambda)^{-1} (\alpha' x + \lambda' \Psi^{-1} y)$$

- The above equation shows that the MIMIC latent factor estimator is a sum of two terms: the first one is the “causes” term (function of) and the second one can be called the “indicators” term.
- If there are no ‘causes' then it reduces to the pure FA estimator.

MIMIC Models (contd.)

Multivariate extension of this model

$$\begin{aligned} y &= \Lambda f + \varepsilon \\ f &= Bx + \varepsilon \end{aligned}$$

$$V(\varepsilon) = \Psi, V(\varepsilon) = \sigma^2 I$$

$$\hat{f} = (I - \Lambda'\Omega^{-1}\Lambda)^{-1}(Bx + \Lambda'\Psi^{-1}y).$$

Structural Equations Models (SEM)

- These are systems of equations involving several latent endogenous variables accounting for the *interdependence* among the latent variables as well as the influence of exogenous “causes”.
- They also have equations describing the ‘measurement’ of the ‘latent’ factors through a set of indicators.

SEM (contd.)

Thus there are two parts:

The structural part explains the latent variables y* (which are the endogenous variables of the model) by a set of exogenous (also latent) variables x* and including mutual effects of the endogenous variables on one another.

$$Ay^{*} + Bx^{*} + u = 0$$

SEM (contd.)

The measurement part specifies the relations linking the latent factors to the observed indicators.

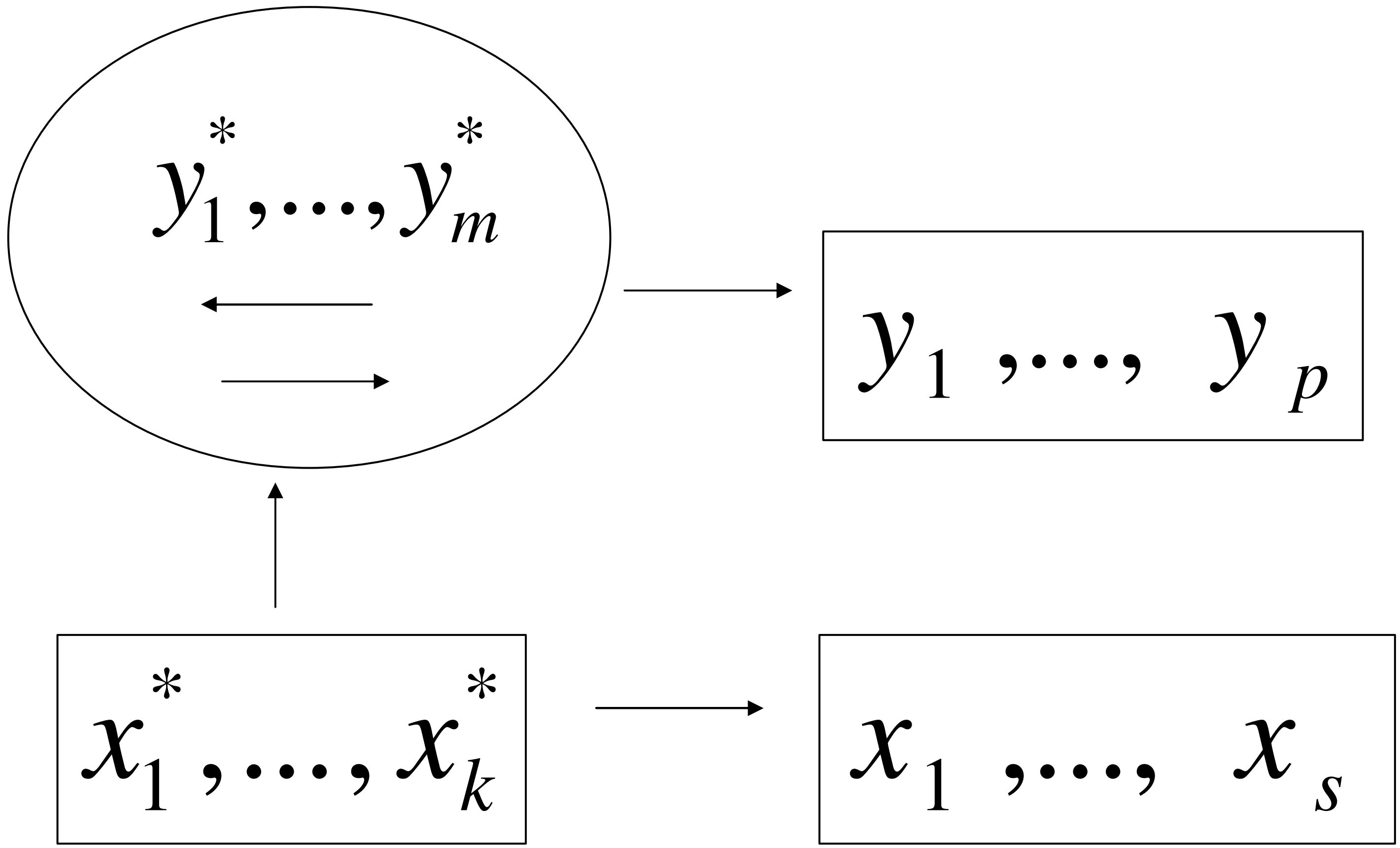
$$y = \Lambda\, y^{*} + \varepsilon$$

$$x = \Upsilon\, x^{*} + \zeta$$

We have

$$V(u) = \Sigma, \quad V(\varepsilon) = \Psi, \quad V(\varepsilon) = \Xi$$

SEM: The Causal Structure



Extension with covariates (exogenous variables)

$$y = \Lambda y^{*} + D w + \varepsilon$$

$$x = \Upsilon x^{*} + F z + \zeta$$

Extension with qualitative outcomes

$$y = h(y^*, w) + \zeta$$

- For a dichotomous indicator , say literate or not, we have:

$$y = 1 \text{ if } y^* > 0 \text{ (say literate) and } y = 0 \text{ if } y^* < 0$$

- For an ordered categorical indicator with C categories (say different levels of education):

$$y = c \text{ if } s_c < y^* < s_{c+1}, \quad c = 1...C, \quad s_1 = -\inf, \; s_C = \inf$$

SEM: Estimation

- The unknown parameters can be estimated by conditional generalised method of moments (GMM) or conditional maximum likelihood (ML).
- Once the parameter estimates are obtained, the latent factors are estimated by their posterior means given the sample, replacing the parameter values by their estimates.

$$\hat{y}_i^* = \left[I - A^{-1}\Sigma A^{-1\prime}\Lambda(\Lambda' A^{-1}\Sigma A^{-1\prime}\Lambda' + \Psi)^{-1}\Lambda \right] A^{-1} B x_i +$$

$$A^{-1}\Sigma A^{-1\prime}\Lambda'(\Lambda A^{-1}\Sigma A^{-1\prime}\Lambda' + \Psi)^{-1} y_i$$