Bayesian approach of structural equation models

S.Demeyer[1,2] N.Fischer[1] G.Saporta[2]

[1]LNE, Laboratoire National de Métrologie et d'Essais

[2]Chaire de Statistique Appliquée & Cedric, CNAM

August 23rd 2010

Le progrès, une passion à partager

PLAN

Example

Situation

The marketing department of a company needs to understand the loyalty of its clients to improve its marketing strategy.
A questionnaire is sent to clients.

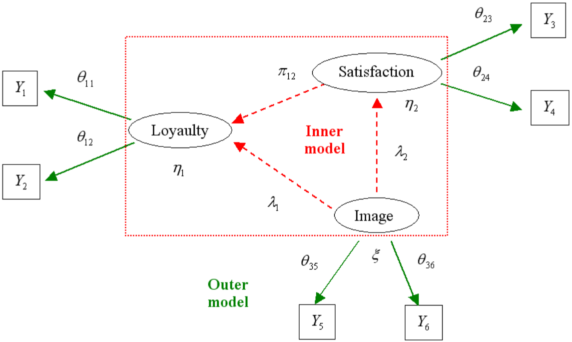
- Direct question : "Are you loyal to our company ?" YES or NO
- More informative : "Would you recommend our company to a friend ?" STRONGLY-MAYBE-NOT AT ALL
- Indirect question : "Are you sensitive to our latest ad ?" or "Are you satisfied with our on-line services ?"

Loyalty is closely related to **satisfaction** and **image**.

Structural Equation Modelling

Answer

Understanding relations between **loyalty**, **satisfaction** and **image** requires Structurel Equation Modeling.



Features of Structural Equation Models

Purpose

To capture latent causality links in data based on expert knowledge.

Features

- multivariate models
- latent variable models
- different types of manifest variables : continuous, binary and ordered categorical

Modelling the example

Equations of the outer model

$$
\begin{aligned}
Y_{i1} &= \theta_{11}\eta_{i1} + E_{i1} \\
Y_{i2} &= \theta_{12}\eta_{i1} + E_{i2} \\
Y_{i3} &= \theta_{23}\eta_{i2} + E_{i3} \\
Y_{i4} &= \theta_{24}\eta_{i2} + E_{i4} \\
Y_{i5} &= \theta_{35}\xi_{i} + E_{i5} \\
Y_{i6} &= \theta_{36}\xi_{i} + E_{i6}
\end{aligned}
$$

Equations of the inner model

$$
\begin{aligned}
\eta_{i1} &= \pi_{12}\eta_{i2} + \lambda_1\xi_i + \delta_{i1} \\
\eta_{i2} &= \lambda_2\xi_i + \delta_{i2}
\end{aligned}
$$

Matricial expression of the example

$$Z = (\eta_1 \, \eta_2 \, \xi) = (H \, \Xi)$$

$$\theta = \begin{pmatrix} \theta_{11} & \theta_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{23} & \theta_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{35} & \theta_{36} \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0 & 0 \\ \pi_{12} & 0 \\ \lambda_1 & \lambda_2 \end{pmatrix}$$

$$(Y_{i1} \dots Y_{i6}) = (Z_{i1} \dots Z_{i3}) \, \theta + E_i$$

$$(H_{i1} \, H_{i2}) = (Z_{i1} \dots Z_{i3}) \, \Lambda + \delta_i$$

Le progrès, une passion à partager

Structural Equation Model

Outer model : $Y_i = Z_i\theta + E_i$
Inner model : simultaneous equation model

$$H_i = Z_i\Lambda + \Delta_i$$

$$\Pi_0^t H_i = \Gamma^t \Xi_i + \Delta_i \qquad \Pi_0 = Id - \Pi$$

Hypotheses

- $E_i \sim \mathcal{N}(0, \Sigma_\varepsilon), \Delta_i \sim \mathcal{N}(0, \Sigma_\delta)$,
- Σ_ε and Σ_δ are diagonal,
- δ_i and Ξ_i are independant,
- Ξ_i is distributed $\mathcal{N}(0, \Phi)$.

Covariance matrix of latent variables

Let $\Sigma_Z = cov(Z)$. Then $\Sigma_Z = \Sigma_Z\left(\pi_{12}, \lambda_1, \lambda_2, \delta_1, \delta_2, \phi\right)$.

Covariance matrix of latent variables in the example

$$
\begin{aligned}
&Var(\eta_1) = \left(\pi_{12}\lambda_2 + \lambda_1\right)^2 \phi + Var\left(\delta_1\right) + \pi_{12}^2 Var\left(\delta_2\right)\\
&Var\left(\eta_2\right) = \lambda_2^2 \phi + Var\left(\delta_2\right)\\
&Var\left(\xi\right) = \phi\\
&Cov\left(\eta_1, \eta_2\right) = \left(\lambda_1 \lambda_2 + \pi_{12}\lambda_2^2\right)\phi + \pi_{12} Var\left(\delta_2\right)\\
&Cov\left(\eta_1, \xi\right) = \left(\pi_{12}\lambda_2 + \lambda_1\right)\phi\\
&Cov\left(\eta_2, \xi\right) = \lambda_2 \phi
\end{aligned}
$$

Formula in terms of the inner parameters

$$
\Sigma_Z = \begin{pmatrix} \left(\Pi_0^t\right)^{-1}\left(\Gamma^t \Phi \Gamma + \Sigma_\delta\right)\Pi_0^{-1} & \left(\Pi_0^t\right)^{-1}\Gamma^t \Phi \\ \Phi \Gamma \Pi_0^{-1} & \Phi \end{pmatrix}
$$

Identifiability issues

Identifiability issues arise because **latent variables are not scaled**. Look at this univariate equation, where Z is a latent variable

$$Y_i = \theta Z_i + E_i$$

The marginal variance of Y is $\theta^2\psi^2 + \sigma^2$ where $\psi^2 = \mathrm{Var}\,(Z)$.
$\tilde{\theta} = \theta\psi$ can be seen as the regression coefficient of the standardized variable $\tilde{Z}$.
So that, equivalently

$$Y_i = \tilde{\theta}\tilde{Z}_i + E_i$$

Le progrès, une passion à partager

Multivariate case

Based on the reduced form distribution [Skrondal,2004]

$$\mathbf{Y_i} \sim N\left(0, \theta^t \Sigma_Z \theta + \Sigma_\varepsilon\right)$$

where Σ_Z is the covariance matrix of the LV.
Yielding the reduced form parameters

$$\psi_{kk}\theta_{ki}^2 + \sigma_{ki}^2,\ i = 1 \ldots n_k,\ k = 1 \ldots K$$
$$\psi_{kk}\theta_{ki}\theta_{kj},\ 1 \leq i < j \leq n_k,\ k = 1 \ldots K$$
$$\psi_{kk'}\theta_{ki}\theta_{k'j},\ 1 \leq i \leq n_k,\ 1 \leq j \leq n_{k'},\ k = 1 \ldots K$$

Identifiability constraints

Three equivalent sets of constraints

	Explained variance		Residual variance
Set 1	$\theta_{k1} = 1$	ψ_{kk} free	$\Sigma_\varepsilon = \text{Id}$
Set 2	θ_{k1} free	$\psi_{kk} = 1$	
Set 3	$\theta_{k1} = 1$	$\psi_{kk} = 1$	Σ_ε free

In the example

$$\psi_{11} = (\pi_{12}\lambda_2 + \lambda_1)^2 \phi + Var(\delta_1) + \pi_{12}^2 Var(\delta_2)$$
$$\psi_{22} = \lambda_2^2 \phi + Var(\delta_2)$$
$$\psi_{33} = \phi$$

Data augmentation

Latent variable are considered as augmented data [Tanner & Wong, 1987].

Thus, latent variables

- have to be estimated as well,
- are conditioned upon to compute posterior distributions of parameters.

Le progrès, une passion à partager

Estimation scheme

A two steps scheme

1. Imputation of latent variables for all statistical units,
2. Estimation of model parameters.

In a bayesian framework

Use of conditional posterior distribution
Use of prior knowledge about parameters
Gibbs algorithm

Le progrès, une passion à partager

Posterior conditional distributions of latent variables

Let $\Theta = \{\theta, \Sigma_\varepsilon, \Lambda, \Sigma_\delta, \Phi\}$ the set of parameters to be estimated.

$$[Z|Y,\Theta] \propto \prod_{i=1}^{n} [Z_i|Y_i,\Theta] \propto \prod_{i=1}^{n} [Y_i|Z_i,\Theta]\,[Z_i|\Theta]$$

- $[Y_i|Z_i,\Theta] = [Y_i|Z_i,\theta,\Sigma_\varepsilon]$
- $[Z_i|\Theta] = [Z_i|\Lambda,\Sigma_\delta,\Phi] \sim \mathcal{N}(0,\Sigma_Z)$ where Σ_Z is the covariance matrix of latent variables.

Results

$$Z_i|Y_i,\theta,\Sigma_\varepsilon,\Lambda,\Sigma_\delta,\Phi \sim \mathcal{N}\left(D\theta\Sigma_\varepsilon^{-1}Y_i, D\right)$$

with $D = \theta\Sigma_\varepsilon^{-1}\theta^t + \Sigma_Z^{-1}$.

Factorization of the posterior distribution of parameters

$$[\Theta|Y,Z] \propto [Y,Z|\Theta]\,[\Theta]$$
$$\propto [Y|Z,\Theta]\,[Z|\Theta]\,[\Theta]$$

- $[Y|Z,\Theta] = [Y|Z,\theta,\Sigma_\varepsilon]$
- $[Z|\Theta] = [Z|\Lambda,\Sigma_\delta,\Phi] = [H|\Xi,\Lambda,\Sigma_\delta]\,[\Xi|\Phi]$
- $[\Theta] = [\theta,\Sigma_\varepsilon]\,[\Lambda,\Sigma_\delta]\,[\Phi]$

$$[\Theta|Y,Z] \quad \propto \quad \underbrace{[Y|Z,\theta,\Sigma_\varepsilon]\,[\theta,\Sigma_\varepsilon]}_{[\theta,\Sigma_\varepsilon|Y,Z]} \; \underbrace{[H|\Xi,\Lambda,\Sigma_\delta]\,[\Lambda,\Sigma_\delta]}_{[\Lambda,\Sigma_\delta|Y,Z]} \; \underbrace{[\Xi|\Phi]\,[\Phi]}_{[\Phi|Z]}$$

Le progrès, une passion à partager

Conjugate prior distributions

$(\theta_k, \Sigma_{\varepsilon k})$ et $(\Lambda_k, \Sigma_{\delta k})$ are Normal-Gamma distributed.

$$\theta_k | \Sigma_{\varepsilon k} \sim \mathcal{N}\left(\theta_{0k}, \Sigma_{\varepsilon k}\Sigma_{\varepsilon 0k}\right)$$
$$(\Sigma_{\varepsilon k})^{-1} \sim \mathcal{G}\left(\alpha_{0\varepsilon k}, \beta_{0\varepsilon k}\right)$$

$$\Lambda_k | \Sigma_{\delta k} \sim \mathcal{N}\left(\Lambda_{0k}, \Sigma_{\delta k}\Sigma_{\delta 0k}\right)$$
$$(\Sigma_{\delta k})^{-1} \sim \mathcal{G}\left(\alpha_{0\delta k}, \beta_{0\delta k}\right)$$

(Ξ, Φ) is Normal-Wishart distributed.

$$\Xi_i | \Phi \sim \mathcal{N}(0, \Phi)$$
$$[\Phi] \propto |\Phi|^{-\frac{1}{2}(q_2+1)}$$

Results

Posterior conditional distribution of $(\theta_k, \Sigma_{\varepsilon k})$

$$\theta_k | Y, Z, \Sigma_\varepsilon \sim \mathcal{N}\left(D_k A_k, \Sigma_{\varepsilon k} D_k\right)$$
$$\Sigma_{\varepsilon k}^{-1} \sim \mathcal{G}\left(\frac{n}{2} + \alpha_{0\varepsilon k}, \beta_{0\varepsilon k} + \frac{1}{2}\left[Y_k^t Y_k - \left(D_k A_k\right)^t D_k^{-1} D_k A_k + \frac{\theta_{0k}^2}{\Sigma_{\varepsilon 0 k}}\right]\right)$$
$$D_k = \left(Z^t Z + \Sigma_{0\varepsilon}^{-1}\right)^{-1}$$
$$A_k = Z^t Y_k + \Sigma_{0\varepsilon}^{-1} \theta_{0k}$$

Le progrès, une passion à partager

Results

Posterior conditional distribution of $(\Lambda_k, \Sigma_{\delta k})$

$$\Lambda_k | Z, \Sigma_{\delta k} \sim \mathcal{N}\left(D_k A_k, \Sigma_{\delta k} D_k\right)$$
$$D_k = \left(Z^t Z + \Sigma_{0\delta}^{-1}\right)^{-1}$$
$$A_k = Z^t H_k + \Sigma_{0\delta}^{-1} \Lambda_{0k}$$

Posterior conditional distribution of Φ

$$\Phi | \Xi \sim \mathcal{IW}\left(\Xi^t \Xi, n\right)$$

Le progrès, une passion à partager

Gibbs algorithm

A step is added to take into account identifiability constraints :

Improved Gibbs algorithm

1. Sample Z from $\mathcal{N}\left(O, \theta^t \Sigma_Z \theta + \Sigma_\varepsilon\right)$,
2. **Set Z^* the standardized latent variables : $Z^* \sim \mathcal{N}\left(0, R_Z\right)$ where R_Z is a correlation matrix,**
3. Sample the parameters conditionnaly to Z^*.

Application : results for the example

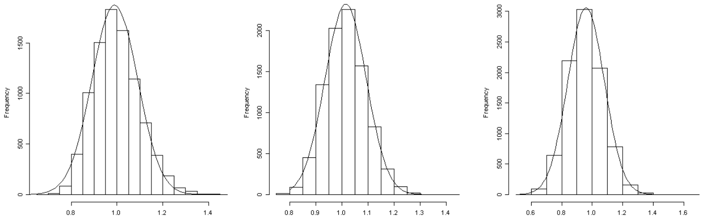
Part of the ECSI model considering only relationships between loyalty, satisfaction and image.

Data from XL Stat : demonstration dataset.

Ordinal variables treated as continuous.

Prior guess about regression coefficients : $\theta_0 = 0.5$, $\Lambda_0 = 0.5$, $\Sigma_{\varepsilon 0} = 1$, $\Sigma_{\delta 0} = 1$ and $\Phi_0 = 1$.

Posterior distributions of constraints

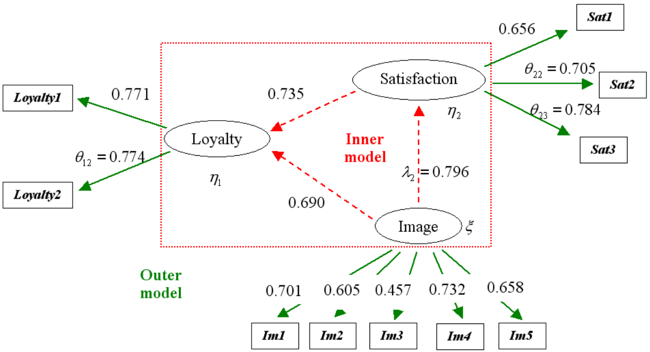


a) $\psi_{11} = (\pi_{12}\lambda_2 + \lambda_1)^2 \phi + Var(\delta_1) + \pi_{12}^2 Var(\delta_2)$

b) $\psi_{22} = \lambda_2^2 \phi + Var(\delta_2)$

c) $\psi_{33} = \Phi$

Results


Sat1
0.656
Satisfaction
$\theta_{22} = 0.705$
Sat2
η_2
$\theta_{23} = 0.784$
Sat3
Loyalty1
0.771
0.735
Loyalty
Inner
model
$\theta_{12} = 0.774$
Loyalty2
η_1
$\lambda_2 = 0.796$
0.690
Image
ξ
Outer
model
0.701
0.605
0.457
0.732
0.658
Im1
Im2
Im3
Im4
Im5

Conclusion and perspectives

Conclusions

- Alternative Bayesian approach to estimate Structural Equation Models,
- The method has been generalized to binary and ordered categorical manifest variables,
- An alternative Gibbs algorithm using **parameter expansion** has been developped to take into account identifiability constraints.

Perspectives

- To compare both Bayesian approaches,
- To compute a univariate score function summarizing the structural part of the model to rate the statistical units.