

15ème Forum des Jeunes Mathématicien-ne-s
du 26 au 28 novembre 2015

Estimation par algorithme EM d'un modèle à facteurs et équation structurelle

Myriam Tami
sous l'encadrement de C. Lavergne, X. Bry

IMAG, UM

- 1 Cadre et problème : modélisation de lien entre groupes de variables
 - Le modèle
 - La question de l'estimation

- 2 Estimation du modèle par EM
 - Généralités sur l'algorithme EM
 - Méthode d'estimation et algorithme

- 3 Application sur données réelles
 - Application : la qualité de vie relative à la santé en cancérologie

- 4 Conclusion

Exemple de modélisation par équations structurelles

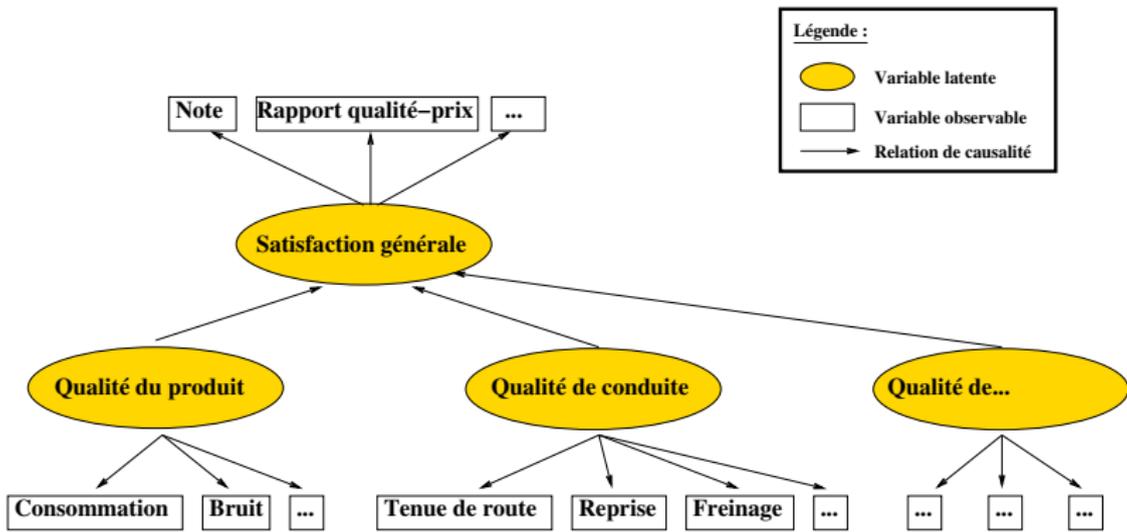


Figure : Étude de la satisfaction d'une gamme automobile : cas de plusieurs variables latentes explicatives et une dépendante

Exemple de modélisation par équations structurelles

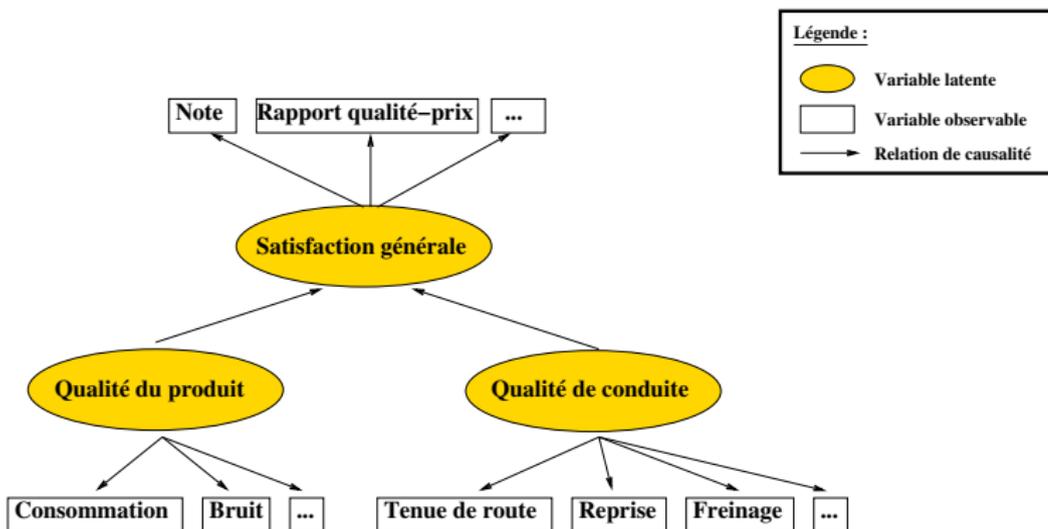


Figure : Étude de la satisfaction d'une gamme automobile : cas de deux variables latentes explicatives et une dépendante

Exemple de modélisation par équations structurelles

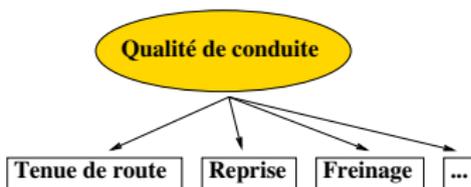
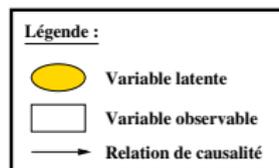


Figure : Étude de la satisfaction d'une gamme automobile : **un modèle de mesure**

Exemple de modélisation par équations structurelles

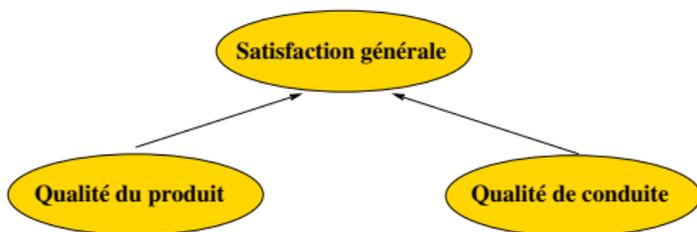
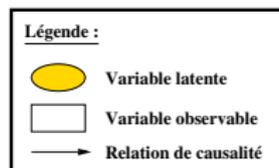


Figure : Étude de la satisfaction d'une gamme automobile : [modèle structurel](#)

Schéma du modèle structurel à 2 facteurs explicatifs et un dépendant

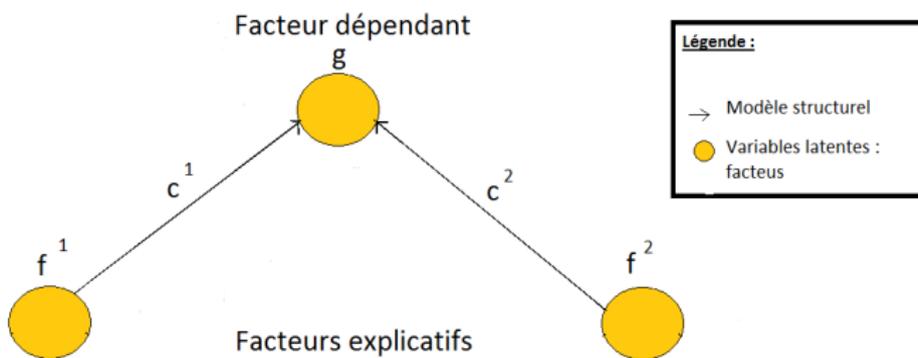


Figure : Modèle structurel à 2 facteurs explicatifs et un dépendant

Écriture du modèle :

$$\left\{ \begin{array}{l} g = f^1 c^1 + \dots + f^2 c^2 + \varepsilon^g \\ \end{array} \right. \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

Schéma du modèle complet à 2 groupes de variables explicatifs et un dépendant

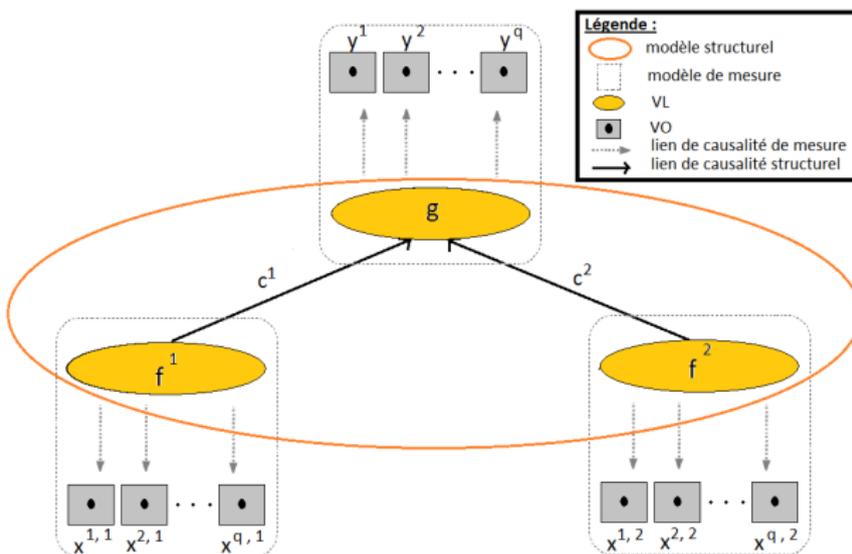


Figure : Modèle complet à 2 groupes de variables explicatifs et un dépendant

Schéma du modèle complet avec adjonction de covariables

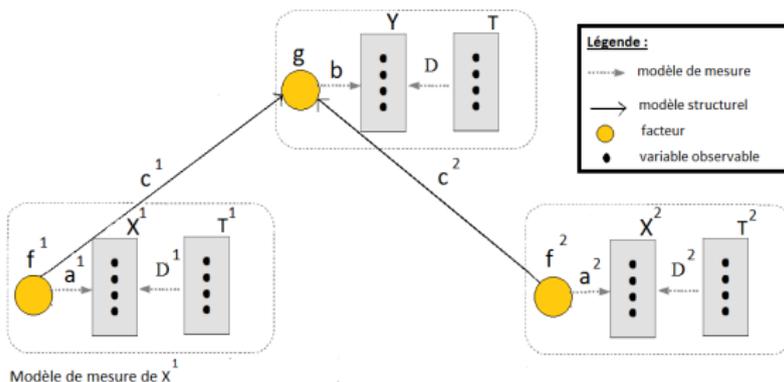


Figure : Modèle complet avec adjonction de covariables

Écriture du modèle :

$$\forall m \in \llbracket 1, 2 \rrbracket \left\{ \begin{array}{l} Y = TD + gb' + \varepsilon^Y \\ X^m = T^m D^m + f^m a^{m'} + \varepsilon^m \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array} \right. \left. \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array} \right.$$

$$\forall m \in \llbracket 1, 2 \rrbracket \left\{ \begin{array}{l} Y = TD + gb' + \varepsilon^Y \\ X^m = T^m D^m + f^m a^{m'} + \varepsilon^m \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array} \right\} \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

Hypothèses

Observations indépendantes $i \in \{1, \dots, n\}$,

- Modèle structurel :

$f^m \sim \mathcal{N}(0, Id_n)$, $m \in \{1, 2\}$ nombre de facteurs explicatifs.

$\varepsilon^g \sim \mathcal{N}(0, Id_n)$

ε^g, f^1, f^2 mutuellement indépendants.

- Modèle de mesure :

$\varepsilon_i^Y \sim \mathcal{N}(0, \sigma_Y^2 Id_{q_Y})$; $\varepsilon_i^m \sim \mathcal{N}(0, \sigma_m^2 Id_{q_m})$

Notation : $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$ de

dimension $K = 5 + q_Y(r_T + 1) + \sum_{m=1}^2 q_m(r_m + 1)$, les paramètres.

Objectif

- 1 Estimation des paramètres θ .
- 2 Estimation des facteurs g, f^m .

↔ Plusieurs approches possibles

Différentes approches : PLS (Wold, 73) vs LISREL (Jöreskog, 1970)

Caractéristiques	LISREL	PLS
Objectif	Validation d'un modèle	Prévision (calcul des Variables Latentes (V.L.))
V.L.	Facteurs non estimés , hypothèse de distribution gaussienne	Composantes calculées (combinaison linéaire de ses Variables Observées (V.O.)) → contrainte
Hypothèses	Indépendance et distribution normale multivariée des données	Indépendance des données
Méthodologie	ASC et Maximum de Vraisemblance (MV)	Estimation des moindres carrés et régressions

↪ Idée : Estimation via **algorithme EM**

Algorithme EM (Dempster et al., 1977)

$\theta \in \Theta$: paramètres

z : les données observées ; h : les données manquantes

$p(\theta)$ fonction de densité de (z, h)

- Estimer $\theta \rightarrow$ maximiser la log-vraisemblance \mathcal{L} complétée des données manquantes.

$$\mathcal{L}(\theta; z, h) = \ln[p(z, h; \theta)]$$

- Algorithme à deux étapes E ("Expectation") et M ("Maximization") qui résout itérativement :

$$\frac{\partial}{\partial \theta} \mathbb{E}_z^h[\mathcal{L}(\theta; z, h)] = 0 \quad (1)$$

Log vraisemblance du modèle

$Z = (Y, X^1, X^2)$ (variables observées)

$h = (g, f^1, f^2)$ (facteurs : nos variables latentes)

La log vraisemblance complétée à maximiser :

$$\begin{aligned} \mathcal{L}(\theta; Z, h) = & -\frac{1}{2} \sum_{i=1}^n \{ \ln|\psi_Y| + \ln|\psi_1| + \ln|\psi_2| \\ & + (y_i - D' t_i - g_i b)' \psi_Y^{-1} (y_i - D' t_i - g_i b) \\ & + (x_i^1 - D^{1'} t_i^1 - f_i^1 a^1)' \psi_1^{-1} (x_i^1 - D^{1'} t_i^1 - f_i^1 a^1) \\ & + (x_i^2 - D^{2'} t_i^2 - f_i^2 a^2)' \psi_2^{-1} (x_i^2 - D^{2'} t_i^2 - f_i^2 a^2) \\ & + (g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2 \} + \lambda \end{aligned}$$

Où, $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$. Avec,

$\psi_Y = \sigma_Y^2 Id_{q_Y}$, $\psi_1 = \sigma_1^2 Id_{q_1}$, $\psi_2 = \sigma_2^2 Id_{q_2}$ et λ une constante.

Résolution : estimation des paramètres

- Par résolution de (1) on obtient

$$\hat{\theta} = \{\hat{D}, \hat{D}^1, \hat{D}^2, \hat{b}, \hat{a}^1, \hat{a}^2, \hat{c}^1, \hat{c}^2, \hat{\sigma}_Y^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2\}$$

Exemples de formules solutions :

$$\hat{c}^1 = \frac{(\sigma_{12} + \overline{f^1 \tilde{g}}) \overline{\tilde{\phi}^2} - (\sigma_{13} + \overline{f^2 \tilde{g}})(\sigma_{23} + \overline{f^1 \tilde{f}^2})}{\overline{\tilde{\phi}^1 \tilde{\phi}^2} - (\sigma_{23} + \overline{f^1 \tilde{f}^2})^2}$$

$$\hat{\sigma}_Y^2 = \frac{1}{nq_Y} \sum_{i=1}^n \{ \|y_i - \hat{D}' t_i\|^2 + \|\hat{b}\|^2 \tilde{\gamma}_i - 2(y_i - \hat{D}' t_i)' \hat{b} \tilde{g}_i \}$$

Résolution : estimation des paramètres

- Par résolution de (1) on obtient

$$\hat{\theta} = \{\hat{D}, \hat{D}^1, \hat{D}^2, \hat{b}, \hat{a}^1, \hat{a}^2, \hat{c}^1, \hat{c}^2, \hat{\sigma}_Y^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2\}$$

Exemples de formules solutions :

$$\hat{c}^1 = \frac{(\sigma_{12} + \overline{f^1 \tilde{g}}) \overline{\phi^2} - (\sigma_{13} + \overline{f^2 \tilde{g}})(\sigma_{23} + \overline{f^1 f^2})}{\overline{\phi^1 \phi^2} - (\sigma_{23} + \overline{f^1 f^2})^2}$$

$$\hat{\sigma}_Y^2 = \frac{1}{nq_Y} \sum_{i=1}^n \{ \|y_i - \widehat{D}' t_i\|^2 + \|\hat{b}\|^2 \tilde{\gamma}_i - 2(y_i - \widehat{D}' t_i)' \hat{b} \tilde{g}_i \}$$

- Les paramètres estimés dépendent des facteurs \tilde{g}_i , \tilde{f}_i^1 , \tilde{f}_i^2 et des carrés des facteurs $\tilde{\gamma}_i$, $\tilde{\phi}_i^1$, $\tilde{\phi}_i^2$.

Résolution : estimation des facteurs

- On explicite :

$$h_i | z_i \sim \mathcal{N}(M_i(\theta) = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i} \end{pmatrix}, \Sigma(\theta) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix})$$

- On estime les facteurs \tilde{g}_i , \tilde{f}_i^1 , \tilde{f}_i^2 et leurs carrés $\tilde{\gamma}_i$, $\tilde{\phi}_i^1$, $\tilde{\phi}_i^2$ par :

$$\tilde{g}_i = \mathbb{E}_{z_i}^{h_i}[g_i] = m_{1i}$$

$$\tilde{f}_i^1 = \mathbb{E}_{z_i}^{h_i}[f_i^1] = m_{2i}$$

$$\tilde{f}_i^2 = \mathbb{E}_{z_i}^{h_i}[f_i^2] = m_{3i}$$

$$\tilde{\gamma}_i = \mathbb{E}_{z_i}^{h_i}[g_i^2] = m_{1i}^2 + \sigma_{11}$$

$$\tilde{\phi}_i^1 = \mathbb{E}_{z_i}^{h_i}[(f_i^1)^2] = m_{2i}^2 + \sigma_{22}$$

$$\tilde{\phi}_i^2 = \mathbb{E}_{z_i}^{h_i}[(f_i^2)^2] = m_{3i}^2 + \sigma_{33}$$

Algorithme

Faire schéma Algo.

Application : étude de la qualité de vie (QdV)

- Données issues d'un questionnaire dont la structure génère la modélisation proposée.

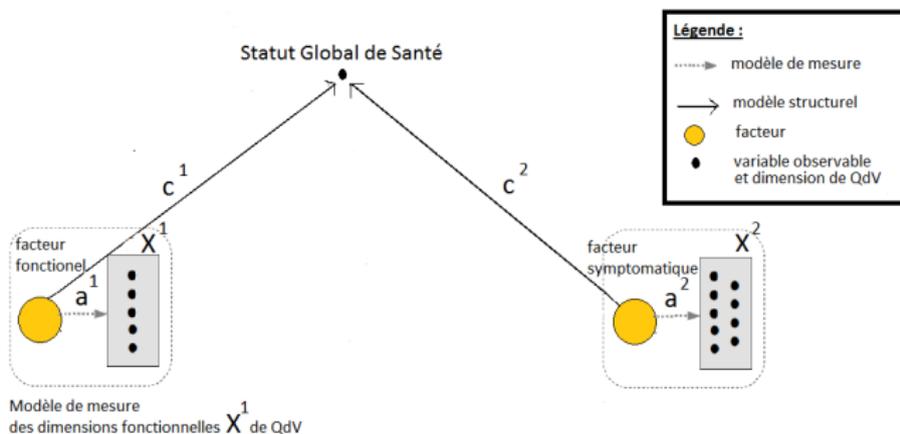


Figure : Modèle conceptuel de la qualité de vie relative à la santé en cancérologie

Proposition d'une approche en deux étapes : protocole et outils

- Objectif : Modéliser le *statut global de santé (QL2)* par un facteur fonctionnel et un facteur symptomatique à plusieurs temps de visite.

	Protocole	Outils
Étape 1	Étude transversale À chaque visite fixée : reconstruction de deux facteurs expliquant le QL2.	SEM, facteurs → 2 facteurs f^1 (resp. f^2) résumant les dimensions fonctionnelles (resp. symptomatiques).
Étape 2	Étude longitudinale Expliquer QL2 à partir des f^1 et f^2 reconstruits à chaque visite.	LMM fondé sur les facteurs f^1, f^2 plus covariables.

- **Avantage** : L'étude longitudinale n'est plus faite dimension par dimension (variable par variable).

Étape 1 : résultats

- Influence du traitement par visite :
 - Corrélation entre \widetilde{f}_R^m avec la covariable R traitement (abscisse) et \widetilde{f}^p sans la covariable (ordonnée) ($m = 1, 2$) aux 8 visites.

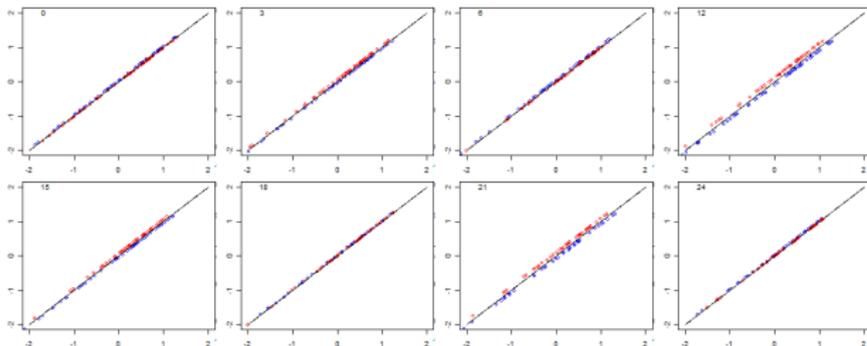


Figure : Corrélations entre \widetilde{f}_R^1 et \widetilde{f}^1 pour 73 patientes traitées par "radiothérapie-létrozole concomitant" et 70 par "radiothérapie suivie par létrozole".

Étude longitudinale (LMM, facteurs)

- Outils : Modèle linéaire mixte (LMM)
- Avantage :
 - Prendre en compte la variabilité induite par les données répétées dans le temps pour un même patient.
 - Quantifier la part d'information apportée par les variables explicatives.

$$y_{iv} = \alpha + \underbrace{x'_{iv}\beta}_{\text{partie fixe}(\widetilde{f'_{iv}} + \text{covariables})} + \underbrace{u'_i\xi_i}_{\text{partie aleatoire}(\text{individu, intercept, visite})} + \varepsilon_{iv}$$

Hypothèses

- $\varepsilon_{iv} \sim \mathcal{N}(0, \sigma^2)$, indépendants des facteurs ;
- $\xi_i \sim \mathcal{N}(\mathbf{0}, \Sigma)$

Étape 2 : analyse longitudinale et choix de modèle (BIC)

$$y_{iv} = \alpha + \beta_1 \tilde{f}_{iv}^1 + \beta_2 \tilde{f}_{iv}^2 + \xi_i + \varepsilon_{iv}$$

- Part d'information portée par les différents éléments du modèle :

	Complet	Sans \tilde{f}_{iv}^2	Sans \tilde{f}_{iv}^1	Sans ξ_i	Sans $(\tilde{f}_{iv}^1, \tilde{f}_{iv}^2)$
BIC	6440.4	6481.6	6516.2	6568.2	6823.0
Différence		41.2	75.8	127.8	382.7

- Le facteur fonctionnel (\tilde{f}_{iv}^1) apporte environ deux fois plus d'information que le facteur symptomatique (\tilde{f}_{iv}^2).
- La part individuelle du modèle (ξ_i) apporte de l'information pour expliquer la variable réponse (QL2).
- Les facteurs sont fortement corrélés (-0.8) \Rightarrow leur présence dans le modèle a du sens.

Conclusion :

Caractéristiques	LISREL	PLS	Estimation par EM
Objectif	Validation d'un modèle	Prévision (calcul des V.L.)	Validation et Prévision
VL	facteurs non estimés	composantes calculées	facteurs estimés
Méthodologie	ASC et max. de vrais.	estimation des moindres carrés et régressions	Max de vrais et EM
Convergence	Matrices non définies positives	observée	les deux
Complexité du modèle possible	modérée (< 100 VO)	grande (100 VL, 1000 VO)	modulable
Taille min. de l'échantillon	grande ($n \in \{200, \dots, 800\}$)	faible ($n \in \{30, \dots, 100\}$)	moyenne $n > 100$
Temps de calcul	quelques minutes	non testé	quelques secondes

MERCI À TOUS POUR VOTRE ATTENTION

Bibliographie

- [1] Kenneth A. Bollen (1989), *Structural Equations With Latent Variables*
- [2] Rivera, P. et Satorra, A. (2002), *Latent Variable and Latent Structure Models*, Marcoulides, G. et Moustaki, I., New Jersey, 85–102.
- [3] Jakobowicz, E. (2007), *Contributions aux modèles d'équations structurelles à variables latentes*, Thèse, Paris, 81–99.
- [4] Bacher, F. (1987), *Les modèles structuraux en psychologie présentation d'un modèle : LISREL*, *Le travail humain*, 347–370.
- [5] Jöreskog, K. (1970), *A general method for analysis of covariance structure*, *Biometrika*.
- [6] Fox, J. (2002), *Structural Equation Models*, <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-sems.pdf>.
- [7] Esposito Vinzi, V. et Trinchera, L. (2008), *Modèles à équations structurelles, approches basée sur les composantes*, URL : http://www.academia.edu/390381/Modeles_a_equations_structurelles_approches_basees_sur_les_composantes, Naples.
- [8] Stan, V. et Saporta, G. (2006), *Une comparaison expérimentale entre les approches PLS et LISREL*, Paris.
- [9] Saidane, M. (2006), *Modèles à facteurs conditionnellement hétéroscédastiques et à structure markovienne cachée pour les séries financières*, Thèse, Montpellier.
- [10] Foulley, J-L. (2002), *Algorithme EM : Theorie et application au modèle mixte*, *Journal de la Société Française de Statistique*, Jouyen-Josas.
- [11] Jourdain, B. (2008), *Probabilités et statistique*, URL : <http://cermics.enpc.fr/~jourdain/probastat/poly.pdf>, Paris.

Résultats moyens sur 100 jeux de données simulés

- Critère d'arrêt : choix de $\varepsilon = 10^{-2}$
- Convergences obtenues ≤ 5 itérations

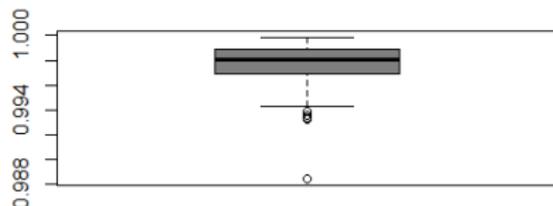


Figure : Boxplot 1

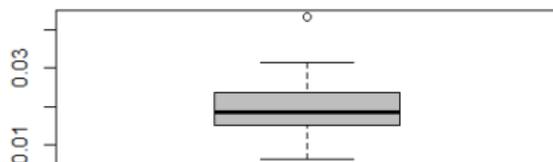


Figure : Boxplot 2

Boxplot 1 : Corrélations entre $(\tilde{g}, \tilde{f}^1, \tilde{f}^2)$ simulés et leurs estimations.

Boxplot 2 : Écarts relatifs en valeur absolue moyens pour chacun des éléments de θ , entre sa valeur simulée et son estimation.

Analyse de sensibilité pour différentes valeurs de q ,

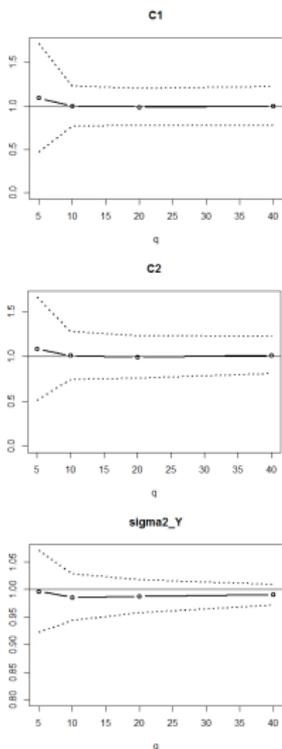


Figure : Moyennes des estimations pour plusieurs valeurs de q des paramètres scalaires c^1 , c^2 , σ_Y^2 avec $IC_{95\%}$

- $q = q_Y = q_1 = q_2 = \{5, 10, 20, 40\}$ pour $n = 400$.
- 100 jeux de données simulés pour chaque cas.
- Mêmes résultats pour σ_1^2 et σ_2^2 .

Analyse de sensibilité pour différentes valeurs de q

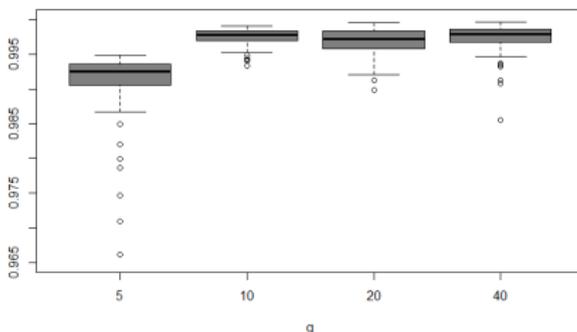


Figure : Boxplot des corrélations entre les facteurs simulés et leur estimations pour différentes valeurs de q

- À partir de $q = 10$, pour un nombre d'observations de $n = 400$ on a une bonne qualité d'estimation des facteurs et des paramètres.

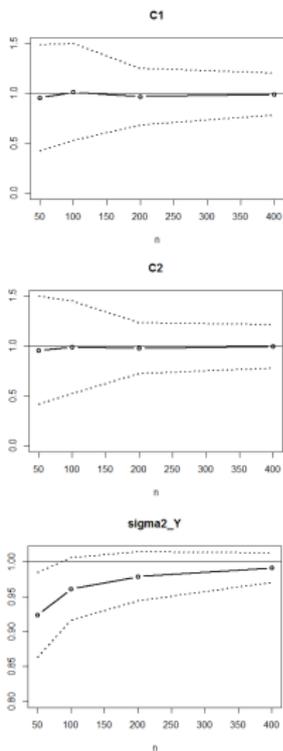
Analyse de sensibilité pour différentes valeurs de n ,

Figure : Moyennes des estimations pour plusieurs valeurs de n des paramètres scalaires c^1 , c^2 , σ_Y^2 avec $IC_{95\%}$

- $n = \{50, 100, 200, 400\}$ pour $q = 40$.
- 100 jeux de données simulés pour chaque cas.
- Mêmes résultats pour σ_1^2 et σ_2^2 .

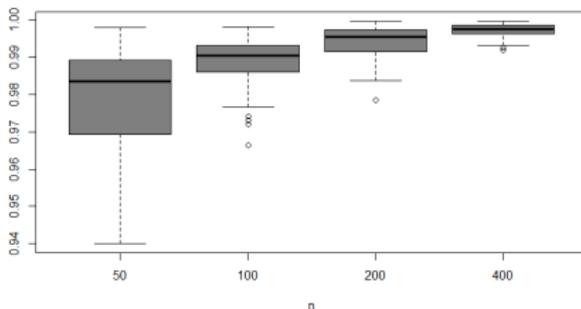
Analyse de sensibilité pour différentes valeurs de n 

Figure : Boxplot des corrélations entre les facteurs simulés et leur estimations pour différentes valeurs de n

- À partir de $n = 100$, pour un nombre de variables de $q = 40$ par bloc on a une bonne qualité d'estimation des facteurs et des paramètres.

Plan de simulation

$$\left\{ \begin{array}{l} Y = TD + gb' + \varepsilon^Y \\ X^1 = T^1 D^1 + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 = T^2 D^2 + f^2 a^{2'} + \varepsilon^{X^2} \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array} \right\} \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

- ① **Choix des dimensions** : $n=400$, $q_Y=q_1=q_2=40$,
 $r_T=r_1=r_2=2$

Plan de simulation

$$\left. \begin{cases} Y = TD + gb' + \varepsilon^Y \\ X^1 = T^1 D^1 + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 = T^2 D^2 + f^2 a^{2'} + \varepsilon^{X^2} \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{cases} \right\} \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

- 1 **Choix des dimensions** : $n=400$, $q_Y=q_1=q_2=40$,
 $r_T=r_1=r_2=2$
- 2 **Définition des paramètres** :

Modèle de mesure

$$D = D^1 = D^2 = (1, 2, \dots, 80)$$

$$b = a^1 = a^2 = (1, 2, \dots, 40)$$

$$\sigma_{X^1}^2 = \sigma_{X^2}^2 = \sigma_Y^2 = 1$$

Plan de simulation

$$\left\{ \begin{array}{l} Y = TD + gb' + \varepsilon^Y \\ X^1 = T^1 D^1 + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 = T^2 D^2 + f^2 a^{2'} + \varepsilon^{X^2} \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array} \right\} \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

- 1 **Choix des dimensions** : $n=400$, $q_Y=q_1=q_2=40$,
 $r_T=r_1=r_2=2$
- 2 **Définition des paramètres** :

Modèle de mesure

$$\begin{aligned} D &= D^1 = D^2 = (1, 2, \dots, 80) \\ b &= a^1 = a^2 = (1, 2, \dots, 40) \\ \sigma_{X^1}^2 &= \sigma_{X^2}^2 = \sigma_Y^2 = 1 \end{aligned}$$

Modèle structurel

$$c^1 = c^2 = 1$$

Plan de simulation

$$\left. \begin{cases} Y = TD + gb' + \varepsilon^Y \\ X^1 = T^1 D^1 + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 = T^2 D^2 + f^2 a^{2'} + \varepsilon^{X^2} \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{cases} \right\} \begin{array}{l} \text{Modèle de mesure} \\ \text{Modèle structurel} \end{array}$$

- 1 **Choix des dimensions** : $n=400$, $q_Y=q_1=q_2=40$,
 $r_T=r_1=r_2=2$
- 2 **Définition des paramètres** :

Modèle de mesure

$$\begin{aligned} D &= D^1 = D^2 = (1, 2, \dots, 80) \\ b &= a^1 = a^2 = (1, 2, \dots, 40) \\ \sigma_{X^1}^2 &= \sigma_{X^2}^2 = \sigma_Y^2 = 1 \end{aligned}$$

Modèle structurel

$$c^1 = c^2 = 1$$

- 3 **Création des données Y, X^1, X^2** :
Simulation de $f^1, f^2, \varepsilon^g \Rightarrow$ création de g .
Simulation de $\varepsilon^Y, \varepsilon^1, \varepsilon^2, T, T^1, T^2 \Rightarrow$ création de Y, X^1, X^2 .

Exemple : l'approche LISREL

Formulation du modèle dans la littérature :

$$\begin{cases} y = \Lambda_y \eta + \epsilon \\ x = \Lambda_x \xi + \delta \quad (\text{Modele de mesure}) \\ \eta = B\eta + \Gamma\xi + \zeta \quad (\text{Modele structurel}) \end{cases}$$

où, Λ_y , Λ_x , B , Γ matrices des coefficients; ϵ , δ , ζ les erreurs de mesure.

Hypothèses :

- ▷ les variables observées sont normales multivariées.
- ▷ ζ et ξ non corrélées.
- ▷ $I - B$ non singulière.
- ▷ $E(\zeta) = E(\xi) = E(\eta) = 0$.
- ▷ $V(\xi) = \Phi$, $V(\zeta) = \Psi$.
- ▷ ϵ et η non corrélées.
- ▷ δ et ξ non corrélées.
- ▷ ϵ , δ et ζ mutuellement non corrélées.
- ▷ $E(\epsilon) = E(\delta) = 0$.
- ▷ $V(\epsilon) = \Theta_\epsilon$, $V(\delta) = \Theta_\delta$.

Exemple : l'approche LISREL

Soit, $\Theta = \{ \Lambda_y, \Lambda_x, B, \Gamma, \Phi, \Psi, \Theta_\epsilon, \Theta_\delta \}$ l'ensemble des paramètres,
 Σ la matrice des covariances entre y et x , supposée définie positive.

- Méthode d'estimation : **hypothèse fondamentale** :

$$\Sigma = \Sigma(\Theta)$$

\Leftrightarrow

$$\begin{pmatrix} V(y) & COV(y, x) \\ COV(x, y) & V(x) \end{pmatrix} = \begin{pmatrix} \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)[(I - B)^{-1}]'\Lambda_y' + \Theta_\epsilon & \Lambda_y(I - B)^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'[(I - B)^{-1}]'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{pmatrix}$$

- **À l'aide de S** la matrice des covariances observées (supposée définie positive), **on explicite $\hat{\Sigma}(\Theta)$ en minimisant la fonction** basée sur le maximum de vraisemblance (ML),

$$F_{ML}(S, \Sigma(\Theta)) = \log|\Sigma(\Theta)| + tr(S\Sigma(\Theta)^{-1}) - \log|S| - cste$$

Autres fonctions à minimiser (moindres carrés) :

$$F_{ULS}(S, \Sigma(\Theta)) = \left(\frac{1}{2}\right)tr((S - \Sigma(\Theta))^2)$$

$$F_{GLS}(S, \Sigma(\Theta)) = \left(\frac{1}{2}\right)tr(S^{-1}(S - \Sigma(\Theta))^2)$$

$$F_{ADF}(S, \Sigma(\Theta)) = (S - \Sigma(\Theta))'W^{-1}(S - \Sigma(\Theta))$$

W une matrice de poids.

L'identité de Fisher (1925) :

$$\frac{\partial}{\partial \theta} \ln[p(z; \theta)] = \mathbb{E}_z^h \left[\frac{\partial}{\partial \theta} \ln[p(z, h; \theta)] \right]$$

- EM revient à résoudre itérativement :

$$\mathbb{E}_z^h \left[\frac{\partial}{\partial \theta} \ln[p(z, h; \theta)] \right] = 0$$

- Une vision classique de EM d'un point de vu algorithmique :

$$E_z^h \left[\frac{\partial}{\partial \theta} \ln[p(z, h; \theta = \theta^{[t]})] \right] = \frac{\partial}{\partial \theta} \left\{ \mathbb{E}_y^f [\ln[p(z, h; \theta = \theta^{[t]})]] \right\}$$

↔ Comme $p(h|z; \theta = \theta^{[t]})$ ne dépend pas de θ , on peut sortir l'opérateur de dérivation de l'espérance conditionnelle à l'itération [t].

$$\begin{aligned}\frac{\partial}{\partial \theta} \ln[p(y; \theta)] = 0 &\iff \frac{1}{p(y; \theta)} \frac{\partial}{\partial \theta} \ln[p(y; \theta)] = 0 \\ &\iff \frac{1}{p(y; \theta)} \frac{\partial}{\partial \theta} \int_f \ln[p(y, f; \theta)] df = 0 \\ &\iff \frac{1}{p(y; \theta)} \int_f \frac{\partial}{\partial \theta} \ln[p(y, f; \theta)] df = 0 \\ &\iff \int_f \frac{\frac{\partial}{\partial \theta} \ln[p(y, f; \theta)]}{p(y, f; \theta)} \frac{p(y, f; \theta)}{p(y; \theta)} df = 0 \\ &\iff \int_f \frac{\partial}{\partial \theta} \ln[p(y, f; \theta)] p(f|y; \theta) df = 0 \\ &\iff \mathbb{E}_y^f \left[\frac{\partial}{\partial \theta} \ln[p(y, f; \theta)] \right] = 0\end{aligned}$$

Où $\mathbb{E}_y^f[\cdot]$ l'espérance par rapport à la distribution conditionnelle des données f sachant y .

Écart relatif à la solution attendue :

$$\frac{|\theta_{estim} - \theta_{simul}|}{\theta_{simul}}$$

$$\widehat{\mu^{X^r}} = \overline{x^r} - \widehat{a^r} \widetilde{f^r}$$

$$\widehat{b} = \frac{\overline{\widetilde{g}y} - \overline{y} \overline{\widetilde{g}}}{\overline{\widetilde{\gamma}} - \overline{\widetilde{g}}^2}$$

$$\widehat{a^r} = \frac{\overline{\widetilde{f^r}x^r} - \overline{x^r} \overline{\widetilde{f^r}}}{\overline{\widetilde{\phi^r}} - \overline{\widetilde{f^r}}^2}$$

$$\widehat{c^1} = \frac{(\sigma_{12} + \overline{\widetilde{f^1} \widetilde{g}}) \overline{\widetilde{\phi^2}} - (\sigma_{13} + \overline{\widetilde{f^2} \widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})}{\overline{\widetilde{\phi^1} \widetilde{\phi^2}} - (\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})^2}$$

$$\widehat{c^2} = \frac{(\sigma_{13} + \overline{\widetilde{f^2} \widetilde{g}}) \overline{\widetilde{\phi^1}} - (\sigma_{12} + \overline{\widetilde{f^1} \widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})}{\overline{\widetilde{\phi^1} \widetilde{\phi^2}} - (\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})^2}$$

$$\widehat{\sigma_{x^r j}^2} = \frac{1}{nq_r} \sum_{i=1}^n \{ \|x_i^r - \widehat{\mu^{X^r}}\|^2 + \|\widehat{a^r}\|^2 \widetilde{\phi}_i^r - 2(x_i^r - \widehat{\mu^{X^r}}) \widehat{a^r} \widetilde{f}_i^r \}$$

Généralisation du modèle à R groupes de variables explicatifs et 1 groupe dépendant

Soit, μ^Y, μ^{X^r} les vecteurs moyennes, A, B matrices des coefficients et $\varepsilon^Y, \varepsilon^{X^r}, \varepsilon^G$ les matrices des erreurs.

Écriture générale du modèle :

$$\begin{cases} Y = \mathbb{1}_n \mu^Y + GB + \varepsilon^Y \\ X^r = \mathbb{1}_n \mu^{X^r} + F^r A^r + \varepsilon^{X^r} \quad \forall r = 1, \dots, R \quad (\text{Modele de mesure}) \\ G = \sum_{r=1}^R \{F^r C^r\} + \varepsilon^G \quad (\text{Modele structurel}) \end{cases}$$

Sous contraintes d'identifiabilité, les **hypothèses** sont :

Soit i une observation,

▷ $F^r \sim \mathcal{N}(0, I)$.

▷ $\varepsilon_i^{X^r} \sim \mathcal{N}(0, \psi_{X^r})$, $\varepsilon_i^Y \sim \mathcal{N}(0, \psi_Y)$.

▷ $\varepsilon_i^G \sim \mathcal{N}(0, I)$.

▷ ε_i^G , mutuellement indépendants des F^r pour tout observation i .

▷ $\psi_{X^r} = \begin{pmatrix} \sigma_{r,1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{r,q_r}^2 \end{pmatrix}$ et $\psi_Y = \begin{pmatrix} \sigma_{Y,1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{Y,q_Y}^2 \end{pmatrix}$