46èmes Journées de Statistique, Rennes, 03 juin 2014

Estimation par algorithme EM de modèles à facteurs et équations structurelles

Myriam Tami
I3M, UM2

X. Bry (I3M), C. Lavergne (I3M)

1. Cadre et problème : modélisation de lien entre groupes de variables
   - Le modèle
   - La question de l'estimation
2. Estimation du modèle par Algorithme EM
   - Généralités sur l'algorithme EM
   - Méthode d'estimation et algorithme
3. Résultats numériques
   - Plan de simulation d'un jeu de données
   - Résultats
4. Conclusion et perspectives

Exemple de modélisation par équations structurelles

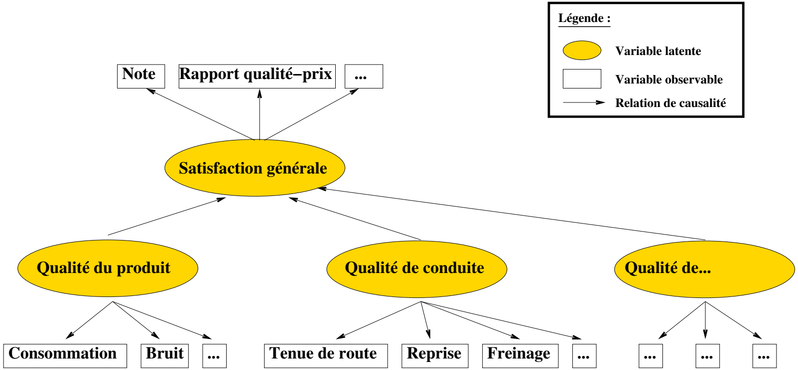

Figure : Étude de la satisfaction d'une gamme automobile : cas de plusieurs variables latentes explicatives et une dépendante

Exemple de modélisation par équations structurelles

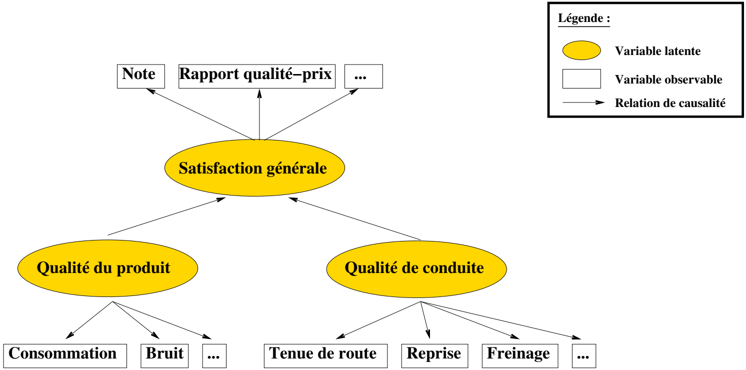

Figure : Étude de la satisfaction d'une gamme automobile : cas de deux variables latentes explicatives et une dépendante

Exemple de modélisation par équations structurelles

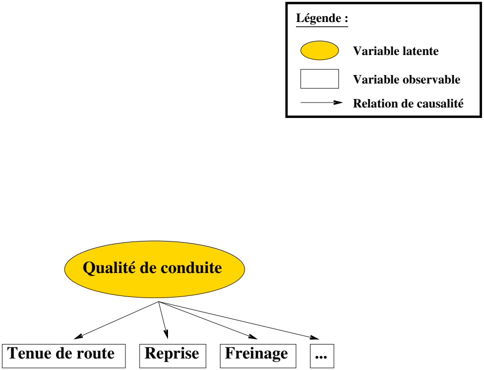

Figure : Étude de la satisfaction d'une gamme automobile : un modèle de mesure

Exemple de modélisation par équations structurelles

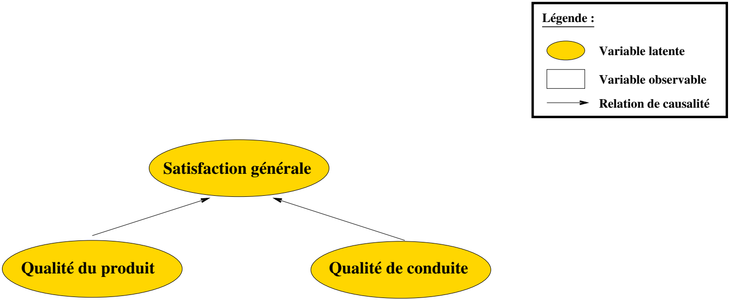

Figure : Étude de la satisfaction d'une gamme automobile : modèle structurel

Schéma du modèle structurel à 2 facteurs explicatifs et un dépendant

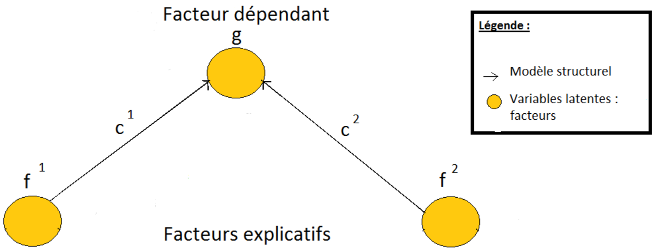

Figure : Modèle structurel à 2 facteurs explicatifs et un dépendant

Écriture du modèle :

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

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

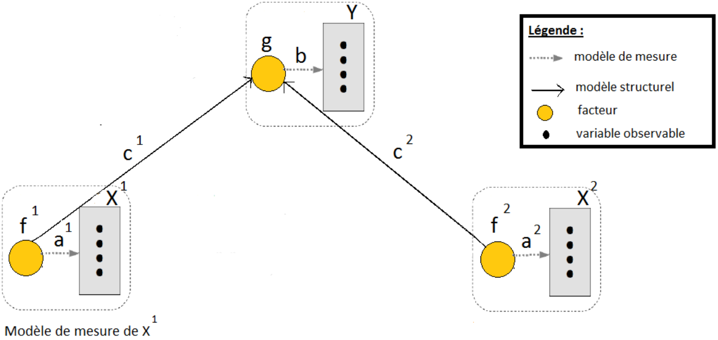

Figure : Modèle complet à 2 groupes de variables explicatifs et un dépendant

Le modèle complet à 2 groupes de variables explicatifs et un dépendant

Écriture du modèle :

$$\left\{\begin{array}{rl} Y = \mathbb{1}_n {\mu^Y}' + g b' + \varepsilon^Y \\ X^1 = \mathbb{1}_n {\mu^{X^1}}' + f^1 {a^1}' + \varepsilon^{X^1} \\ X^2 = \mathbb{1}_n {\mu^{X^2}}' + f^2 {a^2}' + \varepsilon^{X^2} \end{array}\right\} \text{Modèle de mesure}$$

$$g = f^1 c^1 + f^2 c^2 + \varepsilon^g \quad \} \text{ Modèle structurel}$$

Le modèle complet à 2 groupes de variables explicatifs et un dépendant

Modèle et dimensions :

$$
\begin{cases}
Y = \mathbb{1}_n \mu^{Y'} + gb' + \varepsilon^Y & (n, q_Y) = (n,1)(1,q_Y) + (n,1)(1,q_Y) + (n,q_Y) \\
X^1 = \mathbb{1}_n \mu^{X^1{}'} + f^1 a^{1\prime} + \varepsilon^{X^1} & (n, q_1) = (n,1)(1,q_1) + (n,1)(1,q_1) + (n,q_1) \\
X^2 = \mathbb{1}_n \mu^{X^2{}'} + f^2 a^{2\prime} + \varepsilon^{X^2} & (n, q_2) = (n,1)(1,q_2) + (n,1)(1,q_2) + (n,q_2) \\
g = f^1 c^1 + f^2 c^2 + \varepsilon^g & (n,1) = (n,1)(1,1) + (n,1)(1,1) + (n,1)
\end{cases}
$$

n observations.

$$\left\{\begin{array}{rl} Y &= \mathbb{1}_n \mu^{Y'} + gb' + \varepsilon^Y \\ X^1 &= \mathbb{1}_n \mu^{X^{1'}} + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 &= \mathbb{1}_n \mu^{X^{2'}} + f^2 a^{2'} + \varepsilon^{X^2} \\ g &= f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array}\right. \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{Modèle de mesure} \\ \left.\right\} \text{Modèle structurel} \end{array}$$

Hypothèses :
Observations indépendantes $i \in \{1, ..., n\}$,

- Modèle structurel :
  $f^r \sim \mathcal{N}(0, Id_n)$, $r \in \{1, 2\}$ nombre de facteurs explicatifs.
  $\varepsilon^g \sim \mathcal{N}(0, Id_n)$
  ε^g, f^1, f^2 mutuellement indépendants.
- Modèle de mesure :
  $\varepsilon_i^Y \sim \mathcal{N}(0, \sigma_Y^2 Id_{q_Y})$; $\varepsilon_i^{X^r} \sim \mathcal{N}(0, \sigma_{X^r}^2 Id_{q_r})$

Notation : $\theta = \{\mu^Y, \mu^{X^1}, \mu^{X^2}, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_{X^1}^2, \sigma_{X^2}^2\}$, les paramètres.

- But :
  $\rightarrow$ Estimation des paramètres θ.
  $\rightarrow$ Estimation des facteurs g, f^r.

$\hookrightarrow$ Plusieurs approches possibles

Différentes approches

- Approche de type PLS :
  $\hookrightarrow$ PLS (Partial Least Squares) : estimation des moindres carrés (Wold, 1973).
  $\hookrightarrow$ ...
- **Analyse de la structure de covariance** :
  $\hookrightarrow$ LISREL (LInear Structural RELations)(Jöreskog, 1970).
  $\hookrightarrow$ ...
- 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 :

$$
\begin{aligned}
\mathcal{L}(\theta; Z, h) = -\frac{1}{2}\sum_{i=1}^{n}\{ & q_Y ln(\sigma_Y^2) + q_{X^1} ln(\sigma_{X^1}^2) + q_{X^2} ln(\sigma_{X^2}^2) \\
& + \sigma_Y^{-2}(y_i - \mu^Y - g_i b)'(y_i - \mu^Y - g_i b) \\
& + \sigma_{X^1}^{-2}(x_i^1 - \mu^{X^1} - f_i^1 a^1)'(x_i^1 - \mu^{X^1} - f_i^1 a^1) \\
& + \sigma_{X^2}^{-2}(x_i^2 - \mu^{X^2} - f_i^2 a^2)'(x_i^2 - \mu^{X^2} - f_i^2 a^2) \\
& + (g_i - c^1\ f_i^1 - c^2\ f_i^2)^2 \\
& + (f_i^1)^2 + (f_i^2)^2\} + cte
\end{aligned}
$$

Où, $\theta = \{\mu^Y, \mu^{X^1}, \mu^{X^2}, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_{X^1}^2, \sigma_{X^2}^2\}$.

Résolution

- $\hat{\theta} = \{\hat{\mu^Y}, \hat{\mu^{X^1}}, \hat{\mu^{X^2}}, \hat{b}, \hat{a^1}, \hat{a^2}, \hat{c^1}, \hat{c^2}, \hat{\sigma^2_Y}, \hat{\sigma^2_{X^1}}, \hat{\sigma^2_{X^2}}\}$ calculable explicitement par résolution de (1).
  $\hookrightarrow$ Besoin de :
  $h_i|z_i \sim \mathcal{N}(m_i(\theta) = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i} \end{pmatrix}, \Sigma_i(\theta) = \begin{pmatrix} \sigma_{11_i} & \sigma_{12_i} & \sigma_{13_i} \\ \sigma_{21_i} & \sigma_{22_i} & \sigma_{23_i} \\ \sigma_{31_i} & \sigma_{32_i} & \sigma_{33_i} \end{pmatrix})$
- Ex : $\hat{\mu^Y} = \overline{Y} - \widehat{b\overline{\widetilde{g}}}$
  où $\widetilde{g}_i = \mathbb{E}^{h_i}_{z_i}[g_i] = m_{1i}$

Résolution

- $\hat{\theta} = \{\hat{\mu^Y}, \hat{\mu^{X^1}}, \hat{\mu^{X^2}}, \hat{b}, \hat{a^1}, \hat{a^2}, \hat{c^1}, \hat{c^2}, \hat{\sigma^2_Y}, \hat{\sigma^2_{X^1}}, \hat{\sigma^2_{X^2}}\}$ calculable explicitement par résolution de (1).
  $\hookrightarrow$ Besoin de :
  $$h_i|z_i \sim \mathcal{N}(m_i(\theta) = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i} \end{pmatrix}, \Sigma_i(\theta) = \begin{pmatrix} \sigma_{11_i} & \sigma_{12_i} & \sigma_{13_i} \\ \sigma_{21_i} & \sigma_{22_i} & \sigma_{23_i} \\ \sigma_{31_i} & \sigma_{32_i} & \sigma_{33_i} \end{pmatrix})$$
- Ex : $\hat{\mu^Y} = \overline{Y} - \widehat{b\overline{\widetilde{g}}}$
  où $\widetilde{g}_i = \mathbb{E}_{z_i}^{h_i}[g_i] = m_{1i}$
- Les autres paramètres sont aussi calculables et dépendent de :

$$\widetilde{\gamma_i} = \mathbb{E}_{z_i}^{h_i}[g_i^2] = m_{1i}^2 + \sigma_{11_i}$$
$$\widetilde{\phi_i^1} = \mathbb{E}_{z_i}^{h_i}[(f_i^1)^2] = m_{2i}^2 + \sigma_{22_i}$$
$$\widetilde{\phi_i^2} = \mathbb{E}_{z_i}^{h_i}[(f_i^2)^2] = m_{3i}^2 + \sigma_{33_i}$$
$$\widetilde{f_i^1} = \mathbb{E}_{z_i}^{h_i}[f_i^1] = m_{2i}$$
$$\widetilde{f_i^2} = \mathbb{E}_{z_i}^{h_i}[f_i^2] = m_{3i}$$

Algorithme

1. Choix de $\theta^{[0]}$

Algorithme

1. Choix de $\theta^{[0]}$
   Pour tout itération $[t]$,
2. Calcul de la distribution $h|z$.

Algorithme

1. Choix de $\theta^{[0]}$
   Pour tout itération $[t]$,
2. Calcul de la distribution $h|z$.
3. **Étape E** : avec la valeur courante $\theta^{[t]}$,
   $\rightarrow$ Estimation des g, f^r, $\forall r$ via $\widetilde{g} = \mathbb{E}_z^h[g]$, $\widetilde{f}^r = \mathbb{E}_z^h[f^r]$.
   $\rightarrow$ Calcul des $\widetilde{\gamma} = \mathbb{E}_z^h[g^2]$, $\widetilde{\phi}^r = \mathbb{E}_z^h[(f^r)^2]$.

Algorithme

1. Choix de $\theta^{[0]}$
   Pour tout itération $[t]$,
2. Calcul de la distribution $h|z$.
3. **Étape E** : avec la valeur courante $\theta^{[t]}$,
   $\rightarrow$ Estimation des g, f^r, $\forall r$ via $\widetilde{g} = \mathbb{E}_z^h[g]$, $\widetilde{f^r} = \mathbb{E}_z^h[f^r]$.
   $\rightarrow$ Calcul des $\widetilde{\gamma} = \mathbb{E}_z^h[g^2]$, $\widetilde{\phi^r} = \mathbb{E}_z^h[(f^r)^2]$.
4. **Étape M** : on actualise $\theta^{[t+1]}$ en injectant les $\widetilde{g}$, $\widetilde{f^r}$, $\widetilde{\gamma}$, $\widetilde{\phi^r}$ dans les formules solutions de (1).

Algorithme

1. Choix de $\theta^{[0]}$
   Pour tout itération $[t]$,
2. Calcul de la distribution $h|z$.
3. **Étape E** : avec la valeur courante $\theta^{[t]}$,
   $\rightarrow$ Estimation des g, f^r, $\forall r$ via $\widetilde{g} = \mathbb{E}_z^h[g]$, $\widetilde{f^r} = \mathbb{E}_z^h[f^r]$.
   $\rightarrow$ Calcul des $\widetilde{\gamma} = \mathbb{E}_z^h[g^2]$, $\widetilde{\phi^r} = \mathbb{E}_z^h[(f^r)^2]$.
4. **Étape M** : on actualise $\theta^{[t+1]}$ en injectant les $\widetilde{g}$, $\widetilde{f^r}$, $\widetilde{\gamma}$, $\widetilde{\phi^r}$ dans les formules solutions de (1).

On repasse alors à l'étape E puis M, et ainsi de suite.

$$\left\{\begin{array}{l} \left.\begin{array}{r} Y = \mathbb{1}_n {\mu^{Y}}' + gb' + \varepsilon^Y \\ X^1 = \mathbb{1}_n {\mu^{X^1}}' + f^1 {a^1}' + \varepsilon^{X^1} \\ X^2 = \mathbb{1}_n {\mu^{X^2}}' + f^2 {a^2}' + \varepsilon^{X^2} \end{array}\right\} \text{Modèle de mesure} \\ \left.\begin{array}{r} g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array}\right\} \text{Modèle structurel} \end{array}\right.$$

1. **Choix des dimensions** : n=250 ; $q_Y = q_1 = q_2 = 10$

$$\left\{\begin{array}{l} Y = \mathbb{1}_n \mu^{Y'} + g b' + \varepsilon^Y \\ X^1 = \mathbb{1}_n \mu^{X^{1'}} + f^1 a^{1'} + \varepsilon^{X^1} \\ X^2 = \mathbb{1}_n \mu^{X^{2'}} + f^2 a^{2'} + \varepsilon^{X^2} \\ g = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{array}\right. \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{Modèle de mesure} \\ \left.\right\} \text{Modèle structurel} \end{array}$$

1. **Choix des dimensions** : n=250 ; $q_Y = q_1 = q_2 = 10$
2. **Définition des paramètres :**

Modèle de mesure :

$\mu^Y = \mu^{X^1} = \mu^{X^2}$ = vecteur nul de longueur 10

b=(0.07,0.08,0.09,0.1,0.11,0.12,0.13,0.14,0.15,0.2)

$a^1 = a^2$=(0.7,0.8,0.9,1,1.1,1.2,1.3,1.4,0.7,0.8)

$\sigma^2_Y = \sigma^2_{X^1} = \sigma^2_{X^2} = 0.25$

$$\left\{ \begin{array}{l} \left. \begin{array}{l} Y = \mathbb{1}_n {\mu^Y}' + gb' + \varepsilon^Y \\ X^1 = \mathbb{1}_n {\mu^{X^1}}' + f^1 {a^1}' + \varepsilon^{X^1} \\ X^2 = \mathbb{1}_n {\mu^{X^2}}' + f^2 {a^2}' + \varepsilon^{X^2} \end{array} \right\} \text{Modèle de mesure} \\ \left. g = f^1 c^1 + f^2 c^2 + \varepsilon^g \right\} \text{Modèle structurel} \end{array} \right.$$

1. **Choix des dimensions** : n=250 ; $q_Y = q_1 = q_2 = 10$
2. **Définition des paramètres :**

Modèle de mesure :
$\mu^Y = \mu^{X^1} = \mu^{X^2} =$ vecteur nul de longueur 10
b=(0.07,0.08,0.09,0.1,0.11,0.12,0.13,0.14,0.15,0.2)
$a^1 = a^2 =$(0.7,0.8,0.9,1,1.1,1.2,1.3,1.4,0.7,0.8)
$\sigma^2_Y = \sigma^2_{X^1} = \sigma^2_{X^2} = 0.25$

Modèle de structure :
$c^1 = c^2 = 10$

$$\left\{\begin{array}{l} Y = \mathbb{1}_n\mu^{Y'} + gb' + \varepsilon^Y \\ X^1 = \mathbb{1}_n\mu^{X^{1'}} + f^1a^{1'} + \varepsilon^{X^1} \\ X^2 = \mathbb{1}_n\mu^{X^{2'}} + f^2a^{2'} + \varepsilon^{X^2} \\ g = f^1c^1 + f^2c^2 + \varepsilon^g \end{array}\right.$$

Modèle de mesure (les trois premières équations)
Modèle structurel (la dernière équation)

1. **Choix des dimensions** : n=250 ; $q_Y=q_1=q_2=10$
2. **Définition des paramètres :**

Modèle de mesure :
$\mu^Y = \mu^{X^1} = \mu^{X^2} =$ vecteur nul de longueur 10
b=(0.07,0.08,0.09,0.1,0.11,0.12,0.13,0.14,0.15,0.2)
$a^1=a^2$=(0.7,0.8,0.9,1,1.1,1.2,1.3,1.4,0.7,0.8)
$\sigma^2_Y=\sigma^2_{X^1}=\sigma^2_{X^2}=0.25$

Modèle de structure :
$c^1=c^2=10$

3. **Création des données** Y,X^1, X^2 :
Simulation de f^1, f^2 ε^g $\Rightarrow$ création de g.
Simulation de ε^Y, ε^{X^1}, ε^{X^2} $\Rightarrow$ création des données Y, X^1 X^2.

Résultats moyens sur 100 jeux de données

- Écart à la solution attendue : $\|\theta_{attendu} - \theta_{final}\|$

$\hat{\theta}$	$\hat{\mu}^{Y}$	$\hat{\mu}^{X^1}$	$\hat{\mu}^{X^2}$	$\hat{b}$	$\hat{a}^1$	$\hat{a}^2$	$\hat{c}^1$	$\hat{c}^2$	$\hat{\sigma}^2{}_{Y}$, $\hat{\sigma}^2{}_{X^r}$
écart	0.09	0.09	0.09	0.04	0.31	0.33	0.88	0.47	0.03

Conclusion :

- Méthode d'estimation par maximum de vraisemblance via l'algorithme EM.
- Avantage de EM :
  $\rightarrow$ maximisation de log-vaisemblance complexe
  $\rightarrow$ variables latentes

Perspectives :

- Comparer cette méthode aux existantes (efficacité, temps de calcul).
- Généralisation possible

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 markoviene 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.**