

Modèles à Facteurs Conditionnellement
Hétéroscédastiques et à Structure Markovienne
Cachée

Mohamed SAIDANE

18 novembre 2005

Plan Général

1. POURQUOI LES MODÈLES AVEC CHANGEMENT DE RÉGIME ?
2. MODÈLES À FACTEURS STANDARDS ET CHAÎNES DE MARKOV CACHÉES
 - Structure générale du modèle
 - La fonction de vraisemblance
 - L'Algorithme EM
3. MODÈLES À FACTEURS AVEC HÉTÉROSCÉDASTICITÉ DYNAMIQUE ET CHANGEMENT DE RÉGIME
 - Modèle de base et Structure des Facteurs
 - Représentation Espace-État multi-régime et Estimation des Facteurs
 - Algorithme de Filtrage (Méthode GPB1)
 - Algorithme de Lissage
 - L'Algorithme EM
 - L'Approche de Viterbi
4. SIMULATIONS

POURQUOI LES MODÈLES AVEC CHANGEMENT DE RÉGIME ?

Les Modèles à Variables Latentes

- En biostatistique (chaînes de Markov cachées pour le séquençage du génome),
- En traitement d'images (restauration),
- En économie (analyse conjoncturelle, composantes saisonnières),
- En finance (modèles à facteurs, modèles à volatilité stochastique)

Modèles à Facteurs + HMM

- Peut-on distinguer différents régimes caractérisant les rendements financiers ?
- Comment les régimes se diffèrent-ils ?
- Quelle est la fréquence de ces changements de régime et quelles sont leurs dates d'occurrence ?
- Est-ce que le degré des co-mouvements a augmenté ou bien diminué ?
- Les fluctuations communes et spécifiques sont beaucoup ou moins volatiles ?
- Les changements de régime sont-ils prédictibles ?

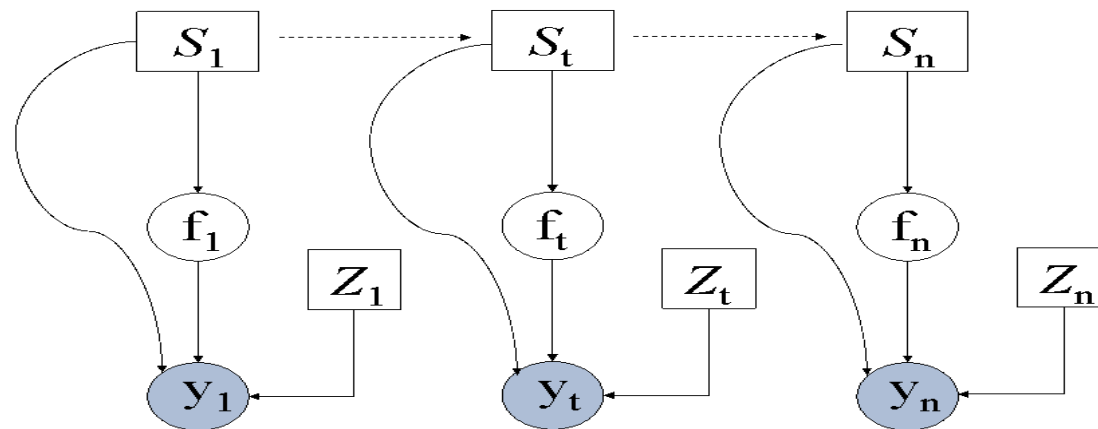
LES MODÈLES FAHMM STANDARDS

I. Structure générale du modèle

$$S_t \sim P(S_t = j / S_{t-1} = i)$$

pour $t = 1, \dots, n$ et $i, j = 1, \dots, m$

$$y_t = \mathbf{X}_{s_t} \mathbf{f}_{s_t} + \varepsilon_{s_t} \quad \text{avec} \quad \begin{cases} \varepsilon_j \sim \mathcal{N}(\theta_j, \Psi_j) \\ \mathbf{f}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_j) \end{cases}$$



II. La fonction de vraisemblance

La vraisemblance d'une observation \mathbf{y}_t étant donné l'état actuel $S_t = j$ peut être obtenue en intégrant par rapport au vecteur d'état \mathbf{f}_t le produit des deux gaussiennes suivantes :

$$\begin{aligned} p(\mathbf{f}_t/S_t = j) &= \mathcal{N}(\mathbf{0}, \mathbf{H}_j) \\ p(\mathbf{y}_t/\mathbf{f}_t, S_t = j) &= \mathcal{N}(\theta_j + \mathbf{X}_j\mathbf{f}_t, \Psi_j) \end{aligned}$$

La vraisemblance résultante est aussi gaussienne et peut être écrite sous la forme :

$$b_j(\mathbf{y}_t) = p(\mathbf{y}_t/S_t = j) = \mathcal{N}(\theta_j, \Sigma_j)$$

où

$$\Sigma_j = \mathbf{X}_j\mathbf{H}_j\mathbf{X}_j' + \Psi_j$$

La vraisemblance complétée d'une séquence d'observations, $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, d'une séquence de vecteurs d'états continus, $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$, et d'une séquence d'états HMM, $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$ est donnée par :

$$p(\mathcal{Y}, \mathcal{F}, \mathcal{S}) = p(S_1) \prod_{t=2}^n p(S_t/S_{t-1}) \prod_{t=1}^n p(\mathbf{f}_t/S_t)p(\mathbf{y}_t/\mathbf{f}_t, S_t)$$

où $p(S_1) = \pi_{s_1}$ est la probabilité de l'état initial et $p(S_t/S_{t-1}) = p_{s_{t-1}s_t}$ sont les probabilités de transition.

III. Estimation des paramètres et l'Algorithme EM

Étape E :

L'espérance conditionnelle de la vraisemblance complétée est donnée par :

$$\begin{aligned} Q(\Theta, \Theta^{(i)}) &= \mathbb{E} \left\{ \log p(\mathcal{Y}, \mathcal{F}, \mathcal{S} / \Theta^{(i)}) / \mathcal{Y}, \Theta \right\} \\ &= \sum_{\forall \mathcal{S}} \int p(\mathcal{F} / \mathcal{Y}, \mathcal{S}, \Theta) p(\mathcal{S} / \mathcal{Y}, \Theta) \log p(\mathcal{Y}, \mathcal{F}, \mathcal{S} / \Theta^{(i)}) d\mathcal{f} \end{aligned}$$

Étape M :

Au niveau de l'étape maximisation, un ensemble de paramètres, $\hat{\Theta}$, maximisant cette espérance conditionnelle sera calculé, soit

$$\hat{\Theta} = \arg \max_{\Theta} Q(\Theta, \hat{\Theta})$$

Ces paramètres seront utilisés pour calculer l'espérance conditionnelle de l'étape E de l'itération suivante, $\hat{\Theta} \longrightarrow \Theta^{(i+1)}$.

1. Les Statistiques a posteriori

1.1. L'Algorithme Arrière-Avant

- Les récurrences α

La probabilité jointe d'une séquence d'observations $\mathcal{Y}_{1:t}$ et de l'état actuel $S_t = j$ est représentée par la variable "Avant" (Forward) $\alpha_j(t) = p(S_t = j, \mathcal{Y}_{1:t})$. Si on suppose que la première observation est générée par le premier état discret, la variable "avant" sera initialisée par :

$$\begin{cases} b_1(\mathbf{y}_1) & , j = 1 \\ 0 & , j \neq 1 \end{cases}$$

Nous pouvons développer la formule de récurrence suivante pour la variable avant à l'instant t :

$$\begin{aligned} \alpha_j(t) &= p(S_t = j, \mathcal{Y}_{1:t}) = p(\mathbf{y}_t / S_t = j) p(S_t = j, \mathcal{Y}_{1:t-1}) \\ &= p(\mathbf{y}_t / S_t = j) \sum_{i=1}^m p(S_t = j, S_{t-1} = i, \mathcal{Y}_{1:t-1}) \\ &= p(\mathbf{y}_t / S_t = j) \sum_{i=1}^m p(S_t = j / S_{t-1} = i) p(S_{t-1} = i, \mathcal{Y}_{1:t-1}) \\ &= b_j(\mathbf{y}_t) \sum_{i=1}^m p_{ij} \alpha_i(t-1) \end{aligned}$$

- Les récurrences β

La probabilité d'une séquence d'observations allant de $t+1$ jusqu'à n conditionnellement à l'état actuel $S_t = j$ est représentée par la variable "Arrière" (Backward), $\beta_i(t) = p(\mathcal{Y}_{t+1:n}/S_t = i)$. Cette variable sera initialisée par $\beta_i(n) = 1 \forall i = 1, \dots, m$.

$$\begin{aligned}\beta_i(t-1) &= p(\mathcal{Y}_{t:n}/S_{t-1} = i) = \sum_{j=1}^m p(S_t = j, \mathcal{Y}_{t:n}/S_{t-1} = i) \\ &= \sum_{j=1}^m p(S_t = j/S_{t-1} = i)p(\mathbf{y}_t/S_t = j)p(\mathcal{Y}_{t+1:n}/S_t = j) \\ &= \sum_{j=1}^m p_{ij}b_j(\mathbf{y}_t)\beta_j(t)\end{aligned}$$

La vraisemblance de la séquence d'observations, \mathcal{Y} , peut être représentée en fonction des variables "arrière" et "avant", soit

$$p(\mathcal{Y}) = \sum_{i=1}^m p(S_t = i, \mathcal{Y}_{1:t})p(\mathcal{Y}_{t+1:n}/S_t = i) = \sum_{i=1}^m \alpha_i(t)\beta_i(t)$$

1.2. Probabilités a posteriori des états discrets

Ces probabilités sont nécessaires pour la mise à jour des paramètres π_j , θ_j , \mathbf{X}_j et Ψ_j .

- Les probabilités γ

$$\begin{aligned}\gamma_j(t) = p(S_t = j/\mathcal{Y}) &= \frac{p(S_t = j, \mathcal{Y})}{p(\mathcal{Y})} \\ &= \frac{p(S_t = j, \mathcal{Y}_{1:t})p(\mathcal{Y}_{t+1:n}/S_t = j)}{p(\mathcal{Y})} \\ &= \frac{\alpha_j(t)\beta_j(t)}{\sum_{i=1}^m \alpha_i(t)\beta_i(t)}\end{aligned}$$

- Les probabilités ξ

Ces probabilités sont nécessaires pour la mise à jour des probabilités de transition p_{ij} .

$$\begin{aligned}\xi_{ij}(t) &= p(S_{t-1} = i, S_t = j/\mathcal{Y}) \\ &= \frac{p(\mathcal{Y}/S_{t-1} = i, S_t = j)p(S_{t-1} = i, S_t = j)}{p(\mathcal{Y})} \\ &= \frac{\alpha_i(t-1)p_{ij}b_j(\mathbf{y}_t)\beta_j(t)}{\sum_{i=1}^m \alpha_i(t)\beta_i(t)}\end{aligned}$$

1.3. Statistiques a posteriori des états continus

Étant donné l'état actuel, $S_t = j$, la distribution jointe du vecteur des observations et du vecteur des facteurs communs est gaussienne, soit

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{f}_t \end{bmatrix} / S_t = j \sim \mathcal{N} \left[\begin{pmatrix} \theta_j \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} \mathbf{X}_j \mathbf{H}_j \mathbf{X}_j' + \Psi_j & \mathbf{X}_j \mathbf{H}_j \\ \mathbf{H}_j \mathbf{X}_j' & \mathbf{H}_j \end{pmatrix} \right]$$

En se basant sur les propriétés de la loi normale multivariée, on démontre que :

$$(\mathbf{f}_t / \mathbf{y}_t, S_t = j) \sim \mathcal{N} \left[\mathbf{K}_j (\mathbf{y}_t - \theta_j), \mathbf{H}_j - \mathbf{K}_j \mathbf{X}_j \mathbf{H}_j \right]$$

où $\mathbf{K}_j = \mathbf{H}_j \mathbf{X}_j' [\mathbf{X}_j \mathbf{H}_j \mathbf{X}_j' + \Psi_j]^{-1}$. On note par

$$\begin{aligned} \hat{\mathbf{f}}_j(t) &= \mathbf{K}_j (\mathbf{y}_t - \theta_j) \\ \hat{\mathbf{R}}_j(t) &= \mathbf{H}_j - \mathbf{K}_j \mathbf{X}_j \mathbf{H}_j + \hat{\mathbf{f}}_j(t) \hat{\mathbf{f}}_j(t)' \end{aligned}$$

2. Mise à jour des paramètres du modèle

2.1. Mise à jour des probabilités initiales

Nous maximisons la fonction Q , soit

$$\begin{cases} Q(\Theta, \Theta^{(i)}) = \sum_{j=1}^m \gamma_j(\mathbf{1}) \log(p(S_1)) \\ \text{S/C } \sum_{j=1}^m \pi_j = 1 \end{cases}$$

où $p(S_1) = \pi = [\pi_1, \pi_2, \dots, \pi_m]'$ et $\pi_j = p(S_1 = j)$.

Si on suppose que la chaîne a commencé à l'état j , l'utilisation du multiplicateur de Lagrange λ avec la contrainte $\sum_{j=1}^m \pi_j = 1$ mène à la maximisation de la fonction :

$$g(\pi_j) = \sum_{i=1}^m \gamma_i(\mathbf{1}) \log(\pi_j) + \lambda \left(1 - \sum_{i=1}^m \pi_i \right)$$

Enfin on trouve :

$$\hat{\pi}_j = \frac{\gamma_j(\mathbf{1})}{\sum_{i=1}^m \gamma_i(\mathbf{1})}$$

2.2. Mise à jour des probabilités de transition

La fonction auxiliaire qu'on cherche à maximiser est donnée par :

$$\left\{ \begin{array}{l} Q(\Theta, \Theta^{(i)}) = \sum_{t=2}^n \sum_{i=1}^m \sum_{j=1}^m \xi_{ij}(t) \log(p_{ij}) \\ \text{S/C } \sum_{j=1}^m p_{ij} = 1 \end{array} \right.$$

Ceci revient à maximiser la fonction :

$$g(p_{ij}) = \lambda \left(1 - \sum_{j=1}^m p_{ij} \right) + \sum_{t=2}^n \sum_{i=1}^m \sum_{j=1}^m \xi_{ij}(t) \log(p_{ij})$$

Les nouvelles probabilités de transition seront données par :

$$\hat{p}_{ij} = \frac{\sum_{t=2}^n \xi_{ij}(t)}{\sum_{t=2}^n \gamma_i(t-1)}$$

2.3. Mise à jour des pondérations

La fonction auxiliaire qu'on cherche à maximiser est donnée par :

$$Q(\Theta, \Theta^{(i)}) = -\frac{1}{2} \sum_{t=1}^n \sum_{j=1}^m \gamma_j(t) \left[\log |\Psi_j| + \mathbb{E} \left\{ \tilde{\mathbf{y}}_{jt}' \Psi_j^{-1} \tilde{\mathbf{y}}_{jt} / \mathcal{Y}, \Theta^{(i)} \right\} \right]$$

où $\tilde{\mathbf{y}}_{jt} = \mathbf{y}_t - \mathbf{X}_j \mathbf{f}_t - \theta_j$. Dans ce cas, si on désigne par \mathbf{x}_{jl} le l ième vecteur ligne de la matrice \mathbf{X}_j , la maximisation de la vraisemblance complétée par rapport à \mathbf{X}_j revient à maximiser la fonction équivalente suivante :

$$g(\mathbf{x}_{jl}) = -\frac{1}{2} \sum_{l=1}^q (\mathbf{x}_{jl} \mathbf{G}_{jl} \mathbf{x}_{jl}' - \mathbf{x}_{jl} \mathbf{k}_{jl})$$

où les matrices \mathbf{G}_{jl} et les vecteurs \mathbf{k}_{jl} sont définis par :

$$\mathbf{G}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^n \gamma_j(t) \hat{\mathbf{R}}_j(t) \quad \text{et} \quad \mathbf{k}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^n \gamma_j(t) (y_{tl} - \theta_{jl}) \hat{\mathbf{f}}_j(t)$$

La résolution des conditions du premier ordre permet de trouver :

$$\hat{\mathbf{x}}_{jl} = \mathbf{k}_{jl}' \mathbf{G}_{jl}^{-1}$$

2.4. Mise à jour des Moyennes

$$\hat{\theta}_j = \frac{\sum_{t=1}^n \gamma_j(t) (\mathbf{y}_t - \mathbf{X}_j \hat{\mathbf{f}}_j(t))}{\sum_{t=1}^n \gamma_j(t)}$$

2.5. Mise à jour des Variances Spécifiques

$$\begin{aligned} \hat{\Psi}_j = & \frac{1}{\sum_{t=1}^n \gamma_j(t)} \sum_{t=1}^n \gamma_j(t) \text{diag} \left\{ \mathbf{y}_t \mathbf{y}_t' - [\mathbf{X}_j \quad \theta_j] \begin{bmatrix} \hat{\mathbf{f}}_j(t) \mathbf{y}_t' \\ \mathbf{y}_t' \end{bmatrix} - [\mathbf{y}_t \hat{\mathbf{f}}_j(t)' \quad \mathbf{y}_t] \begin{bmatrix} \mathbf{X}_j' \\ \theta_j' \end{bmatrix} \right. \\ & \left. + [\mathbf{X}_j \quad \theta_j] \begin{bmatrix} \hat{\mathbf{R}}_j(t) & \hat{\mathbf{f}}_j(t) \\ \hat{\mathbf{f}}_j(t)' & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_j' \\ \theta_j' \end{bmatrix} \right\} \end{aligned}$$

2.6. Mise à jour des Variances des Facteurs

$$\hat{\mathbf{H}}_j = \frac{1}{\sum_{t=1}^n \gamma_j(t)} \text{diag} \left\{ \sum_{t=1}^n \gamma_j(t) \hat{\mathbf{R}}_j(t) \right\}$$

MODÈLES À FACTEURS AVEC HÉTÉROSCÉDASTICITÉ DYNAMIQUE ET CHANGEMENT DE RÉGIME

I. Modèle de base et Structure des Facteurs

$$S_t \sim P(S_t = j / S_{t-1} = i)$$
$$\mathbf{f}_{s_t} = \mathbf{H}_{s_t}^{1/2} \mathbf{f}_t^* \quad \text{où} \quad \mathbf{f}_t^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$$
$$\mathbf{y}_t = \mathbf{X}_{s_t} \mathbf{f}_{s_t} + \varepsilon_{s_t} \quad \text{avec} \quad \varepsilon_{s_t} \sim \mathcal{N}(\theta_{s_t}, \Psi_{s_t})$$

pour $t = 1, \dots, n$ et $i, j = 1, \dots, m$

Les facteurs communs suivent des processus **GQARCH(1,1)** :

$$h_{is_t} = w_{is_t} + \gamma_{is_t} f_{is_{t-1}} + \alpha_{is_t} f_{is_{t-1}}^2 + \delta_{is_t} h_{is_{t-1}}$$

$$h_{is_t} = w_{is_t} + \Theta_{is_t} (f_{is_{t-1}} - \mu_{is_t})^2 + \delta_{is_t} h_{is_{t-1}}$$

- si $\gamma_{is_t} = 0$: une représentation GARCH(1,1)
- si en plus $\delta_{is_t} = 0$: une représentation ARCH(1)
- si $\gamma_{is_t} = \delta_{is_t} = \alpha_{is_t} = 0$: le cas FAHMM standard

II. Représentation Espace-État et Estimation des Facteurs : La Méthode Pseudo-Bayésienne Généralisée

1. Notations

Il faut tout d'abord définir quelques notations, soient

$$\begin{aligned}\mathbf{f}_{t/\tau}^{i(j)} &= \mathbb{E} [\mathbf{f}_t / \mathcal{Y}_{1:\tau}, S_{t-1} = i, S_t = j] \\ \mathbf{f}_{t/\tau}^{(j)k} &= \mathbb{E} [\mathbf{f}_t / \mathcal{Y}_{1:\tau}, S_t = j, S_{t+1} = k] \\ \mathbf{f}_{t/\tau}^j &= \mathbb{E} [\mathbf{f}_t / \mathcal{Y}_{1:\tau}, S_t = j]\end{aligned}$$

Si $\tau = t$, ces dernières seront appelées **statistiques de filtrage** ; si $\tau > t$, on les appelle **statistiques de lissage** ; et si $\tau < t$ on les appelle **statistiques de prédiction**. L'indexe entre parenthèses est la valeur du noeud de changement à la date t ; le terme à gauche est la valeur de S_{t-1} , et celui qui se trouve à droite, représente la valeur de S_{t+1} .

$$\begin{aligned}h_{lt/\tau}^j &= \text{Var} (f_{lt} / \mathcal{Y}_{1:\tau}, S_t = j) \\ h_{lt,t-1/\tau}^j &= \text{Cov} (f_{lt}, f_{lt-1} / \mathcal{Y}_{1:\tau}, S_t = j) \\ h_{lt,t-1/\tau}^{i(j)} &= \text{Cov} (f_{lt}, f_{lt-1} / \mathcal{Y}_{1:\tau}, S_{t-1} = i, S_t = j) \\ M_{t-1,t/\tau}(i, j) &= p (S_{t-1} = i, S_t = j / \mathcal{Y}_{1:\tau}) \\ M_{t/\tau}(j) &= p(S_t = j / \mathcal{Y}_{1:\tau}) \\ L_t(i, j) &= p(\mathbf{y}_t / \mathcal{Y}_{1:t-1}, S_{t-1} = i, S_t = j)\end{aligned}$$

où $L_t(i, j)$ est la vraisemblance de l'innovation à l'instant t , étant donné que le système est à l'état $S_t = j$ sachant que $S_{t-1} = i$.

2. L'Algorithme de Filtrage

2.1. Les équations de mesure et de transition

$$\mathbf{y}_t = \theta_{s_t} + \mathbf{X}_{s_t} \mathbf{f}_{s_t} + \varepsilon_{s_t} \quad \text{et} \quad \mathbf{f}_{s_t} = \mathbf{0} \cdot \mathbf{f}_{s_{t-1}} + \mathbf{f}_{s_t}$$

2.2. Les équations de prévision

Nous effectuons les opérations suivantes successivement :

$$\begin{aligned} \mathbf{f}_{t/t-1}^{i(j)} &= \mathbf{0} \cdot \mathbf{f}_{t-1/t-1}^i = \mathbf{0} \quad \forall i, j = 1, \dots, m \quad \text{et} \\ h_{lt/t-1}^{i(j)} &= w_{lj} + \gamma_{lj} f_{lt-1/t-1}^i + \alpha_{lj} \left[f_{lt-1/t-1}^{i2} + h_{lt-1/t-1}^i \right] + \delta_{lj} h_{lt-1/t-2}^i \\ \mathbf{H}_{t/t-1}^{i(j)} &= \text{diag} \left[h_{lt/t-1}^{i(j)} \right] \quad \text{avec} \quad l = 1, 2, \dots, k \end{aligned}$$

Nous calculons par la suite l'erreur de prédiction, la variance de l'erreur, la matrice de gain de Kalman, et la vraisemblance de cette innovation :

$$\mathbf{e}_t(i, j) = \mathbf{y}_t - \theta_j - \mathbf{X}_j \mathbf{f}_{t/t-1}^{i(j)}$$

$$\Sigma_{t/t-1}^{i(j)} = \mathbf{X}_j \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' + \Psi_j$$

$$\mathbf{K}_t(i, j) = \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' \Sigma_{t/t-1}^{i(j)-1}$$

$$L_t(i, j) = \mathcal{N} \left[\mathbf{0}, \Sigma_{t/t-1}^{i(j)} \right]$$

2.3. Les équations de mise à jour

Les équations de mise à jour pour la moyenne et la variance sont données par :

$$\begin{aligned}\mathbf{f}_{t/t}^{i(j)} &= \mathbf{f}_{t/t-1}^{i(j)} + K_t(i, j)\mathbf{e}_t(i, j) \\ \mathbf{H}_{t/t}^{i(j)} &= \left[\mathbf{I}_k - K_t(i, j)\mathbf{X}_j \right] \mathbf{H}_{t/t-1}^{i(j)} = \mathbf{H}_{t/t-1}^{i(j)} - K_t(i, j)\Sigma_{t/t-1}^{i(j)}K_t(i, j)'\end{aligned}$$

Nous calculons aussi les probabilités

$$M_{t-1,t/t}(i, j) = p(S_{t-1} = i, S_t = j / \mathcal{Y}_{1:t}) = \frac{L_t(i, j)p_{ij}M_{t-1/t-1}(i)}{\sum_{i=1}^m \sum_{j=1}^m L_t(i, j)p_{i,j}M_{t-1/t-1}(i)}$$

Finalement, nous calculons les probabilités

$$\begin{aligned}M_{t/t}(j) &= \sum_{i=1}^m M_{t-1,t/t}(i, j) \\ Z_{i/j}(t) &= p(S_{t-1} = i / S_t = j, \mathcal{Y}_{1:t}) = \frac{M_{t-1,t/t}(i, j)}{M_{t/t}(j)}\end{aligned}$$

2.4. Fusion des Moments Conditionnels

$$\mathbf{f}_{t/t}^j = \sum_{i=1}^m Z_{i/j}(t) \mathbf{f}_{t/t}^{i(j)}$$

$$h_{lt/t}^j = \sum_{i=1}^m Z_{i/j}(t) h_{lt/t}^{i(j)} + \sum_{i=1}^m Z_{i/j}(t) \left[f_{lt/t}^{i(j)} - f_{lt/t}^j \right] \left[f_{lt/t}^{i(j)} - f_{lt/t}^j \right]'$$

$$h_{lt/t-1}^j = \sum_{i=1}^m Z_{i/j}(t) h_{lt/t-1}^{i(j)} + \sum_{i=1}^m Z_{i/j}(t) \left[f_{lt/t-1}^{i(j)} - f_{lt/t-1}^j \right] \left[f_{lt/t-1}^{i(j)} - f_{lt/t-1}^j \right]'$$

$$\mathbf{H}_{t/t}^j = \text{diag} \left[h_{lt/t}^j \right] \quad \text{et} \quad \mathbf{H}_{t/t-1}^j = \text{diag} \left[h_{lt/t-1}^j \right] \quad \text{pour} \quad l = 1, 2, \dots, k$$

3. L'Algorithme de Lissage

Étant donnée la nature dégénérée de l'équation de transition, la matrice de gain de lissage est toujours nulle, soit

$$J_t^{(j)k} = \mathbf{H}_{t/t}^j \mathbf{0}'_k \mathbf{H}_{t+1/t}^{(j)k-1} = \mathbf{0}$$

ce qui implique :

$$\begin{aligned} \mathbf{f}_{t/n}^{(j)k} &= \mathbf{f}_{t/t}^j + J_t^{(j)k} \left[\mathbf{f}_{t+1/n}^k - \mathbf{f}_{t+1/t}^{j(k)} \right] = \mathbf{f}_{t/t}^j \\ \mathbf{H}_{t/n}^{(j)k} &= \mathbf{H}_{t/t}^j + J_t^{(j)k} \left[\mathbf{H}_{t+1/n}^k - \mathbf{H}_{t+1/t}^{j(k)} \right] J_t^{(j)k'} = \mathbf{H}_{t/t}^j \end{aligned}$$

Nous calculons par la suite les probabilités,

$$U_{t/t+1}^{j/k} = p(S_t = j / S_{t+1} = k, \mathcal{Y}_{1:n}) \simeq \frac{M_{t/t}(j) p_{jk}}{\sum_{j'=1}^m M_{t/t}(j') p_{j'k}}$$

où l'approximation provient du fait que S_t n'est pas conditionnellement indépendante du futur $\mathbf{y}_{t+1}, \dots, \mathbf{y}_n$ étant donné l'état S_{t+1} .

Pour la mise à jour des paramètres, nous avons besoin aussi des probabilités

$$M_{t,t+1/n}(j, k) = U_{t/t+1}^{j/k} M_{t+1/n}(k) \quad \text{et} \quad M_{t/n}(j) = \sum_{k=1}^m M_{t,t+1/n}(j, k)$$

III. L'Algorithme EM

La log-vraisemblance complétée

La vraisemblance complétée d'une séquence d'observations, $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, d'une séquence de vecteurs d'états continus, $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$, et d'une séquence d'états HMM, $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$ est donnée par :

$$p(\mathcal{Y}, \mathcal{F}, \mathcal{S}) = p(S_1) \prod_{t=2}^n p(S_t/S_{t-1}) \prod_{t=1}^n p(\mathbf{f}_t/S_t, \mathcal{D}_{1:t-1}) p(\mathbf{y}_t/\mathbf{f}_t, S_t, \mathcal{D}_{1:t-1})$$

où $\mathcal{D}_{1:t-1} = \{\mathcal{Y}_{1:t-1}, \mathcal{F}_{1:t-1}, \mathcal{S}_{1:t-1}\}$, est l'ensemble d'informations disponibles à la date $t-1$, $p(S_1) = \pi_{s_1}$ et $p(S_t/S_{t-1}) = p_{s_{t-1}s_t}$.

Étape E :

Soient $\mathcal{D}_n^{(i)} = \{\mathcal{Y}_{1:n}, \Theta^{(i)}\}$ et $\tilde{\mathbf{y}}_{jt} = \mathbf{y}_t - \mathbf{X}_j \mathbf{f}_t^j$. L'espérance conditionnelle de la log-vraisemblance complétée est donnée par :

$$\begin{aligned} Q(\Theta/\Theta^{(i)}) &\simeq \sum_{j=1}^m M_{1/n}(j) \log p(S_1) - \sum_{t=2}^n \sum_{i=1}^m \sum_{j=1}^m M_{t-1,t/n}(i, j) \log p_{ij} \\ &- \frac{1}{2} \sum_{j=1}^m \sum_{t=1}^n M_{t/n}(j) \left[\log |\Psi_j| + \mathbb{E} \left\{ (\tilde{\mathbf{y}}_{jt} - \theta_j)' \Psi_j^{-1} (\tilde{\mathbf{y}}_{jt} - \theta_j) / \mathcal{D}_n^{(i)} \right\} \right] \\ &- \frac{1}{2} \sum_{j=1}^m \sum_{l=1}^k \sum_{t=1}^n M_{t/n}(j) \mathbb{E} \left[\log(h_{lt}^j) + \frac{f_{lt}^2}{h_{lt}^j} / \mathcal{D}_n^{(i)} \right] \end{aligned}$$

Étape CM1 :

- Comme dans le cas des modèles FAHMM on a :

$$\hat{\pi}_j = \frac{M_{1/n}(j)}{\sum_{i=1}^m M_{1/n}(i)} \quad \text{et} \quad \hat{p}_{ij} = \frac{\sum_{t=2}^n M_{t-1,t/n}(i,j)}{\sum_{t=2}^n M_{t-1/n}(i)}$$

- De même, pour

$$\mathbf{G}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^n M_{t/n}(j) \left[\mathbf{H}_{t/n}^j + \mathbf{f}_{t/n}^j \mathbf{f}_{t/n}^{j'} \right] \quad \text{et} \quad \mathbf{k}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^n M_{t/n}(j) (y_{tl} - \theta_{jl}) \mathbf{f}_{t/n}^j$$

Les pondérations seront données par :

$$\hat{\mathbf{x}}_{jl} = \mathbf{k}'_{jl} \mathbf{G}_{jl}^{-1}$$

- Les estimations des moyennes sont données par :

$$\hat{\theta}_j = \frac{\sum_{t=1}^n M_{t/n}(j) (\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{t/n}^j)}{\sum_{t=1}^n M_{t/n}(j)}$$

- Les variances spécifiques seront données par :

$$\hat{\Psi}_j = \frac{1}{\sum_{t=1}^n M_{t/n}(j)} \sum_{t=1}^n M_{t/n}(j) \text{diag} \left\{ \mathbf{y}_t \mathbf{y}_t' - \begin{bmatrix} \mathbf{X}_j & \theta_j \end{bmatrix} \begin{bmatrix} \mathbf{f}_{t/n}^j \mathbf{y}_t' \\ \mathbf{y}_t' \end{bmatrix} - \begin{bmatrix} \mathbf{y}_t \mathbf{f}_{t/n}^{j'} & \mathbf{y}_t \end{bmatrix} \begin{bmatrix} \mathbf{X}_j' \\ \theta_j' \end{bmatrix} \right. \\ \left. + \begin{bmatrix} \mathbf{X}_j & \theta_j \end{bmatrix} \begin{bmatrix} \left(\mathbf{H}_{t/n}^j + \mathbf{f}_{t/n}^j \mathbf{f}_{t/n}^{j'} \right) & \mathbf{f}_{t/n}^j \\ \mathbf{f}_{t/n}^{j'} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_j' \\ \theta_j' \end{bmatrix} \right\}$$

Étape CM 2 :

La log-vraisemblance non complétée est donnée par :

$$\mathcal{L}^* = c - \frac{1}{2} \sum_{t=1}^n \sum_{j=1}^m S_t(j) \left[\log |\Sigma_{t/t-1}^{(j)}| + (\mathbf{y}_t - \theta_j)' \Sigma_{t/t-1}^{(j)-1} (\mathbf{y}_t - \theta_j) \right]$$

Les dérivées premières de cette fonction par rapport aux paramètres $\phi_{lj} = \{w_{lj}, \gamma_{lj}, \alpha_{lj}, \delta_{lj}\}$, $j = 1, \dots, m$ et $l = 1, 2, \dots, k$, sont données par :

$$\frac{\partial \mathcal{L}^*(\Theta/\mathcal{Y})}{\partial \phi_j} = -\frac{1}{2} \sum_{t=1}^n \frac{\partial \mathbf{h}_{t/t-1}^{(s_t)'}}{\partial \phi_j} \text{vecd} \left[\mathbf{X}_{s_t}' \Sigma_{t/t-1}^{(s_t)-1} \left[\Sigma_{t/t-1}^{(s_t)} - (\mathbf{y}_t - \theta_{s_t})(\mathbf{y}_t - \theta_{s_t})' \right] \Sigma_{t/t-1}^{(s_t)-1} \mathbf{X}_{s_t} \right]$$

Dérivées de la variance conditionnelle

pour $l = 1, 2, \dots, k$

$$\frac{\partial h_{lt/t-1}^{(s_t)}}{\partial w_{lj}} = 1 \quad \text{si } S_t = j \text{ et } 0 \text{ sinon}$$

$$\frac{\partial h_{lt/t-1}^{(s_t)}}{\partial \alpha_{lj}} = \left[f_{lt-1/t-1}^{(s_{t-1})2} + h_{lt-1/t-1}^{(s_{t-1})} \right] + \delta_{lj} \frac{\partial h_{lt-1/t-2}^{(s_{t-1})}}{\partial \alpha_{lj}}$$

$$\frac{\partial h_{lt/t-1}^{(s_t)}}{\partial \gamma_{lj}} = f_{lt-1/t-1}^{(s_{t-1})} + \delta_{lj} \frac{\partial h_{lt-1/t-2}^{(s_{t-1})}}{\partial \gamma_{lj}}$$

$$\frac{\partial h_{lt/t-1}^{(s_t)}}{\partial \delta_{lj}} = h_{lt-1/t-2}^{(s_{t-1})} + \delta_{lj} \frac{\partial h_{lt-1/t-2}^{(s_{t-1})}}{\partial \delta_{lj}}$$

L'algorithme itératif utilisé pour la maximisation de la log-vraisemblance non complétée \mathcal{L}^* est donné par la formule suivante :

$$\phi_j^{(i+1)} = \phi_j^{(i)} + \left[H_{(i)} \left(\phi_j^{(i)} \right) \right]^{-1} g_{(i)} \left(\phi_j^{(i)} \right)$$

où $\phi_j^{(i)}$ est le vecteur contenant les paramètres de la i ème itération ; $H_{(i)} \left(\phi_j^{(i)} \right)$ est une approximation de la matrice Hessienne de \mathcal{L}^* par rapport aux paramètres, évaluée à $\phi_j^{(i)}$; et $g_{(i)} \left(\phi_j^{(i)} \right)$ est le gradient négatif de \mathcal{L}^* évalué à $\phi_j^{(i)}$.

IV. Inférence des Structures Cachées et Approximation de Viterbi

Cette approche est basée sur l'approximation du modèle par un modèle espace-état linéaire par morceau et la minimisation du coût Hamiltonien :

$$\begin{aligned}
 \mathcal{H}(\mathcal{F}_n, \mathcal{S}_n, \mathcal{Y}_n) &= \frac{n}{2}(m + q) \log 2\pi \\
 &+ \frac{1}{2} \sum_{t=1}^n \sum_{j=1}^m \left[(\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{jt} - \theta_j)' \boldsymbol{\Psi}_j^{-1} (\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{jt} - \theta_j) + \log |\boldsymbol{\Psi}_j| \right] S_t(j) \\
 &+ \frac{1}{2} \sum_{t=1}^n \sum_{j=1}^m \left[\mathbf{f}_{jt}' \mathbf{H}_{jt}^{-1} \mathbf{f}_{jt} + \log |\mathbf{H}_{jt}| \right] S_t(j) + \sum_{t=2}^n S'_t(-\log \mathbf{P}) S_{t-1} + S'_1(-\log \pi_1)
 \end{aligned}$$

Le "meilleur" coût partiel jusqu'à la date t de la séquence des observations \mathcal{Y}_t lorsque le système est à l'état j à la date t sera défini par :

$$J_{t,j} = \min_{\mathcal{S}_{t-1}, \mathcal{F}_t} \mathcal{H} \left[\mathcal{F}_t, \{\mathcal{S}_{t-1}, S_t = j\}, \mathcal{Y}_t \right]$$

Pour une transition quelconque $i \rightarrow j$ on a :

$$\begin{aligned} \mathbf{f}_{t/t-1}^{i(j)} &= \mathbf{0} & \forall i, j = 1, \dots, m \quad \text{et} \\ h_{lt/t-1}^{i(j)} &= w_{lj} + \gamma_{lj} f_{lt-1/t-1}^i + \alpha_{lj} (f_{lt-1/t-1}^{i2} + h_{lt-1/t-1}^i) + \delta_{lj} h_{lt-1/t-2}^i \\ \mathbf{H}_{t/t-1}^{i(j)} &= \text{diag} \left[h_{lt/t-1}^{i(j)} \right] & \text{pour } l = 1, 2, \dots, k \end{aligned}$$

Lorsque une nouvelle observation \mathbf{y}_t devient disponible à la date t , ces estimations de prédiction seront filtrées

$$\begin{aligned} \mathbf{f}_{t/t}^{i(j)} &= \mathbf{f}_{t/t-1}^{i(j)} + K_{i,j} \left[\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{t/t-1}^{i(j)} - \theta_j \right] \\ \mathbf{H}_{t/t}^{i(j)} &= [\mathbf{I}_k - K_{i,j} \mathbf{X}_j] \mathbf{H}_{t/t-1}^{i(j)} \end{aligned}$$

Le coût d'innovation $J_{t,t-1,i,j}$ de chaque transition $i \rightarrow j$ est donné par :

$$\begin{aligned} J_{t,t-1,i,j} &= \frac{1}{2} (\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{t/t-1}^{i(j)} - \theta_j)' \left[\mathbf{X}_j \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' + \Psi_j \right]^{-1} (\mathbf{y}_t - \mathbf{X}_j \mathbf{f}_{t/t-1}^{i(j)} - \theta_j) \\ &+ \frac{1}{2} \log |\mathbf{X}_j \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' + \Psi_j| - \log p_{ij} \end{aligned}$$

Une partie de ce coût d'innovation reflète la transition de l'état continu (les facteurs), soient les termes d'innovation de l'équation précédente. Le coût restant, $-\log p_{ij}$, est dû à la transition HMM de l'état i à l'état j .

Pour chaque état actuel j , il y a m états possibles qui peuvent être à son origine. Afin de minimiser le coût total à chaque instant t et pour chacun des états j , on doit sélectionner le "meilleur" état précédent i , soit

$$J_{t,j} = \min_i \{J_{t,t-1,i,j} + J_{t-1,i}\}$$

$$\delta_{t-1,j} = \arg \min_i \{J_{t,t-1,i,j} + J_{t-1,i}\}$$

L'indexe de cet état sera récupéré dans la variable $\delta_{t-1,j}$. Ainsi, les m meilleures estimations de filtrage des états du modèle à facteurs et leurs variances à la date t seront données par :

$$\mathbf{f}_{t/t}^j = \mathbf{f}_{t/t}^{\delta_{t-1,j}(j)} \quad \text{et} \quad \mathbf{H}_{t/t}^j = \mathbf{H}_{t/t}^{\delta_{t-1,j}(j)} \quad \text{avec} \quad h_{lt/t-1}^j = h_{lt/t-1}^{\delta_{t-1,j}(j)}$$

Une fois toutes les n observations $\mathcal{Y}_{1:n}$ seront traitées, le meilleur coût global sera obtenu par :

$$J_n^* = \min_j J_{n,j}$$

Pour décoder la séquence d'états de transition optimale, on utilise l'indexe du meilleur état final, $j_n^* = \arg \min_j J_{n,j}$, et par la suite on procède d'une manière rétrograde à travers la variable qui contient tout l'historique des meilleurs états de transition $\delta_{t-1,j}$,

$$j_t^* = \delta_{t,j_{t+1}^*}$$

Algorithme de VITERBI

Initialisation des statistiques de prédiction $\mathbf{f}_{0/-1}^j$ et $\mathbf{H}_{0/-1}^j$; et du coût $J_{0,j}$.

for $t = 1 : n$

for $j = 1 : m$

for $i = 1 : m$

 Algorithme de filtrage de kalman

 Calcul de $\mathbf{f}_{t/t}^{i(j)}$ et $\mathbf{H}_{t/t}^{i(j)}$

 Calculer le coût d'innovation $J_{t/t-1,i,j}$

end

 Calculer le "meilleur" coût partiel $J_{t,j}$, l'état de transition $\delta_{t-1,j}$, et les estimations des statistiques $\mathbf{f}_{t/t}^j$ et $\mathbf{H}_{t/t}^j$.

end

end

Identification du "meilleur" état de transition j_n^* .

Récurrances arrières \rightarrow la "meilleure" séquence d'états, j_t^* .

Calculer les statistiques exhaustives du modèle.

SIMULATIONS

- Exactitude et Stabilité des Estimations
- Normalité Asymptotique des Estimations
- Sélection de Modèles

I. Simulations

1. Exactitude et Stabilité des Estimations

Dans la première expérience nous nous intéressons au comportement des estimations du FAHMM avec hétéroscédasticité dynamique lorsque la taille de la séquence n augmente de 800 à 1500.

. Les vrais paramètres du modèle, (.) Paramètres d'initialisation de l'algorithme EM

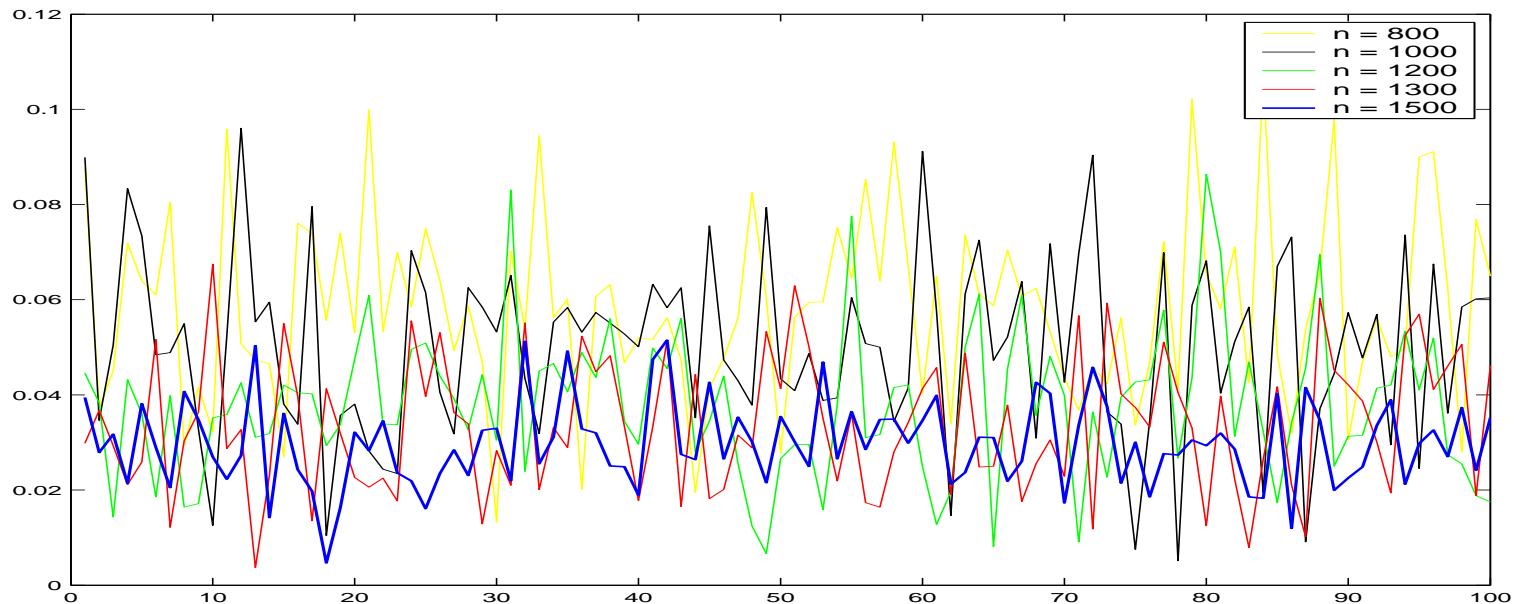
	θ	X	$diag(\Psi)$	ϕ
État 1	1.0000 (0.0000)	2.0000 (1.0000)	1.0000 (0.5000)	0.1000 (0.1000)
	2.0000 (1.0000)	2.0000 (1.0000)	1.0000 (0.5000)	0.3000 (0.1000)
	1.0000 (0.5000)	2.0000 (0.5000)	1.0000 (0.5000)	0.4000 (0.1000)
	2.0000 (1.0000)	2.0000 (4.0000)	1.0000 (0.5000)	
	1.0000 (0.0000)	2.0000 (4.0000)	1.0000 (0.5000)	
	2.0000 (0.5000)	2.0000 (1.0000)	1.0000 (0.5000)	
État 2	2.0000 (1.0000)	3.0000 (1.0000)	2.0000 (0.5000)	0.2000 (0.1000)
	3.0000 (1.0000)	3.0000 (0.5000)	2.0000 (0.5000)	0.2000 (0.1000)
	2.0000 (1.0000)	3.0000 (0.5000)	2.0000 (0.5000)	0.6000 (0.1000)
	3.0000 (1.0000)	3.0000 (1.0000)	2.0000 (0.5000)	
	2.0000 (1.0000)	3.0000 (0.5000)	2.0000 (0.5000)	
	3.0000 (1.0000)	3.0000 (0.5000)	2.0000 (0.5000)	

La métrique naturelle permettant de mesurer la distance des estimateurs par rapport aux vrais paramètres du modèle est la divergence de Kullback-Leibler,

$$K(\Theta_0, \Theta) = \lim_n \frac{1}{n} \left\{ \log \mathcal{L}(y_1, \dots, y_n; \Theta_0) - \log \mathcal{L}(y_1, \dots, y_n; \Theta) \right\}$$

où Θ_0 est l'ensemble des vrais paramètres. Pour une séquence finie de longueur n , la divergence de Kullback-Leibler empirique entre deux ensembles de paramètres sera définie par la formule :

$$K_n(\Theta_0, \Theta) = \frac{1}{n} \left\{ \log \mathcal{L}(y_1, \dots, y_n; \Theta_0) - \log \mathcal{L}(y_1, \dots, y_n; \Theta) \right\}$$



Simulation with $n = 1000$

	θ	X	$diag(\Psi)$	ϕ
State 1	0.8820 (0.1128)	2.0277 (0.0522)	0.9978 (0.0611)	0.0664 (0.0912)
	1.8938 (0.1085)	2.0264 (0.0443)	0.9975 (0.0622)	0.2994 (0.0571)
	0.8817 (0.1165)	2.0255 (0.0427)	0.9994 (0.0593)	0.3936 (0.0510)
	1.8866 (0.1132)	2.0257 (0.0449)	1.0014 (0.0681)	
	0.8865 (0.0960)	2.0313 (0.0386)	1.0076 (0.0644)	
	1.8894 (0.1059)	2.0269 (0.0472)	1.0026 (0.0669)	
State 2	1.8600 (0.1567)	3.0602 (0.0359)	2.0150 (0.1269)	0.1914 (0.1082)
	2.8530 (0.1671)	3.0619 (0.0371)	1.9775 (0.1125)	0.2063 (0.0397)
	1.8570 (0.1633)	3.0617 (0.0388)	1.9956 (0.0983)	0.5817 (0.0374)
	2.8587 (0.1685)	3.0589 (0.0322)	1.9878 (0.1243)	
	1.8537 (0.1684)	3.0608 (0.0366)	1.9818 (0.1382)	
	2.8517 (0.1592)	3.0611 (0.0374)	1.9893 (0.1176)	
Simulation with $n = 1500$				
State 1	0.8969 (0.0771)	2.0232 (0.0307)	0.9955 (0.0472)	0.1030 (0.0585)
	1.8977 (0.0795)	2.0223 (0.0289)	0.9988 (0.0408)	0.2986 (0.0467)
	0.8966 (0.0759)	2.0219 (0.0263)	0.9948 (0.0487)	0.3917 (0.0377)
	1.8999 (0.0913)	2.0221 (0.0288)	0.9968 (0.0397)	
	0.8981 (0.0854)	2.0214 (0.0255)	1.0020 (0.0371)	
	1.8962 (0.0766)	2.0219 (0.0314)	0.9975 (0.0416)	
State 2	1.8315 (0.1389)	3.0579 (0.0211)	1.9949 (0.0774)	0.1826 (0.0755)
	2.8359 (0.1271)	3.0568 (0.0297)	1.9878 (0.0782)	0.1952 (0.0287)
	1.8309 (0.1285)	3.0554 (0.0331)	1.9905 (0.0931)	0.5921 (0.0194)
	2.8349 (0.1194)	3.0592 (0.0266)	1.9847 (0.0763)	
	1.8296 (0.1117)	3.0581 (0.0182)	1.9872 (0.0861)	
	2.8316 (0.1308)	3.0596 (0.0286)	2.0022 (0.0965)	

Simulation with $n = 1200$
The Generalized Pseudo-Bayesian Method

	θ	X	$diag(\Psi)$	ϕ
State 1	0.8832 (0.0931)	2.0254 (0.0417)	0.9993 (0.0483)	0.0695 (0.0901)
	1.8905 (0.0911)	2.0249 (0.0410)	1.0019 (0.0461)	0.3039 (0.0522)
	0.8828 (0.0783)	2.0227 (0.0285)	0.9882 (0.0572)	0.3872 (0.0389)
	1.8900 (0.1326)	2.0216 (0.0357)	0.9960 (0.0626)	
	0.8839 (0.0875)	2.0219 (0.0271)	0.9983 (0.0594)	
	1.8851 (0.0774)	2.0227 (0.0369)	1.0007 (0.0477)	
State 2	1.8581 (0.1521)	3.0569 (0.0317)	1.9980 (0.1014)	0.1884 (0.0782)
	2.8617 (0.1476)	3.0532 (0.0306)	2.0040 (0.0887)	0.2013 (0.0366)
	1.8539 (0.1493)	3.0534 (0.0388)	1.9788 (0.0957)	0.5890 (0.0324)
	2.8584 (0.1505)	3.0527 (0.0287)	1.9776 (0.1059)	
	1.8573 (0.1514)	3.0590 (0.0293)	2.0039 (0.1062)	
	2.8547 (0.1386)	3.0540 (0.0312)	1.9986 (0.1107)	
The Viterbi Approximation				
State 1	0.9839 (0.0983)	1.9980 (0.0682)	0.9989 (0.0472)	0.1017 (0.0865)
	1.9912 (0.0974)	1.9973 (0.0667)	1.0016 (0.0457)	0.3022 (0.0483)
	0.8934 (0.0857)	1.9952 (0.0589)	0.9878 (0.0593)	0.3977 (0.0366)
	1.9906 (0.0861)	1.9940 (0.0571)	0.9958 (0.0607)	
	0.9845 (0.0973)	1.9945 (0.0577)	0.9980 (0.0572)	
	1.9856 (0.0866)	1.9952 (0.0604)	1.0006 (0.0486)	
State 2	1.9726 (0.1833)	3.0134 (0.0475)	1.9988 (0.0984)	0.2046 (0.0876)
	2.9759 (0.1872)	3.0097 (0.0481)	2.0047 (0.0961)	0.1992 (0.0377)
	1.9681 (0.1867)	3.0099 (0.0463)	1.9797 (0.1092)	0.5982 (0.0281)
	2.9726 (0.1954)	3.0092 (0.0454)	1.9783 (0.1067)	
	1.9718 (0.1988)	3.0154 (0.0508)	2.0046 (0.1089)	
	2.9690 (0.1866)	3.0105 (0.0511)	1.9992 (0.1112)	

2. Normalité Asymptotique des Estimations

Nous avons utilisé la statistique de Shapiro-Wilk pour tester la normalité univariée des composantes de $\tilde{\Theta}_n$.

Il s'agit de tester H_0 : les Θ_i suivent la loi $\mathcal{N}(m, \sigma)$, avec m et σ inconnus. Les procédures de ce test sont les suivantes :

- On calcule la statistique \mathcal{W} ,

$$\mathcal{W} = \frac{\left[\sum_{i=1}^r a_{v+1-i} (\Theta_{(v+1-i)} - \Theta_{(i)}) \right]^2}{\sum_{i=1}^v (\Theta_{(i)} - \bar{\Theta})^2}$$

où $r = (v - 1)/2$, si v est impaire, et $r = v/2$ si v est paire ; v c'est le nombre des replications. La distribution exacte de \mathcal{W} sous l'hypothèse nulle dépend seulement de v et non pas des vraies valeurs de m et σ .

- Si \mathcal{W} est plus petite que la valeur critique pour la valeur appropriée de α , on rejette H_0 à un niveau α .

Simulation with $n = 800$ (left values)
 Simulation with $n = 1000$ (right values)

	θ	X	$diag(\Psi)$	ϕ
State 1	0.9792 - 0.9735	0.9676 - 0.9784	0.9797 - 0.9653	0.9734 - 0.9692
	0.9725 - 0.9807	0.9643 - 0.9748	0.9700 - 0.9576	0.9526 - 0.9866
	0.9720 - 0.9784	0.9831 - 0.9857	0.9768 - 0.9877	0.9798 - 0.9831
	0.9713 - 0.9669	0.9763 - 0.9779	0.9586 - 0.9690	
	0.9842 - 0.9797	0.9825 - 0.9608	0.9728 - 0.9710	
	0.9805 - 0.9773	0.9693 - 0.9890	0.9777 - 0.9822	
State 2	0.9738 - 0.9666	0.9519 - 0.9691	0.9802 - 0.9592	0.9809 - 0.9810
	0.9759 - 0.9801	0.9823 - 0.9776	0.9714 - 0.9820	0.9832 - 0.9699
	0.9912 - 0.9914	0.9481 - 0.9584	0.9768 - 0.9662	0.9845 - 0.9397
	0.9749 - 0.9858	0.9686 - 0.9790	0.9605 - 0.9709	
	0.9758 - 0.9765	0.9588 - 0.9745	0.9769 - 0.9857	
	0.9842 - 0.9891	0.9837 - 0.9771	0.9705 - 0.9661	

Significance level α								
Lower tail								Upper tail
0.01	0.02	0.05	0.10	0.50	0.10	0.05	0.02	0.01
0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

3. Sélection de Modèles

$$\checkmark \text{ AIC}(\mathcal{M}) = -2 \mathcal{L}(\hat{\Theta}/\mathcal{Y}) + 2 v_{\mathcal{M}}$$

$$\checkmark \text{ BIC}(\mathcal{M}) = -2 \mathcal{L}(\hat{\Theta}/\mathcal{Y}) + v_{\mathcal{M}} \log n$$

$$\checkmark \text{ ICL}(\mathcal{M}) = \text{BIC}(\mathcal{M}) + \log p(\mathcal{Z}|\mathcal{Y}, \hat{\Theta}_{\mathcal{M}})$$

$\log p(\mathcal{Z}|\mathcal{Y}, \hat{\Theta}_{\mathcal{M}})$ étant une mesure de l'information manquante portée par le modèle \mathcal{M} . Cette écriture met en évidence que le critère ICL sur-pénalise les modèles à information manquante importante par rapport à BIC.

Tel qu'il est défini, le critère ICL n'est pas calculable puisque les états \mathcal{Z} sont non observés. Une approximation naturelle pour $\log p(\mathcal{Z}|\mathcal{Y}, \hat{\Theta}_{\mathcal{M}})$ est donnée par :

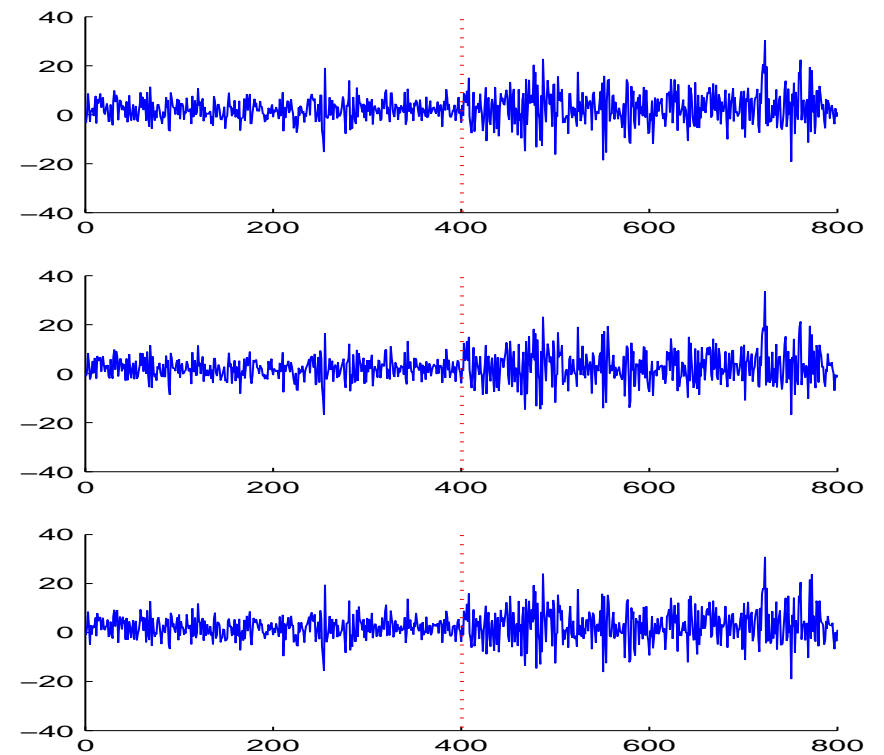
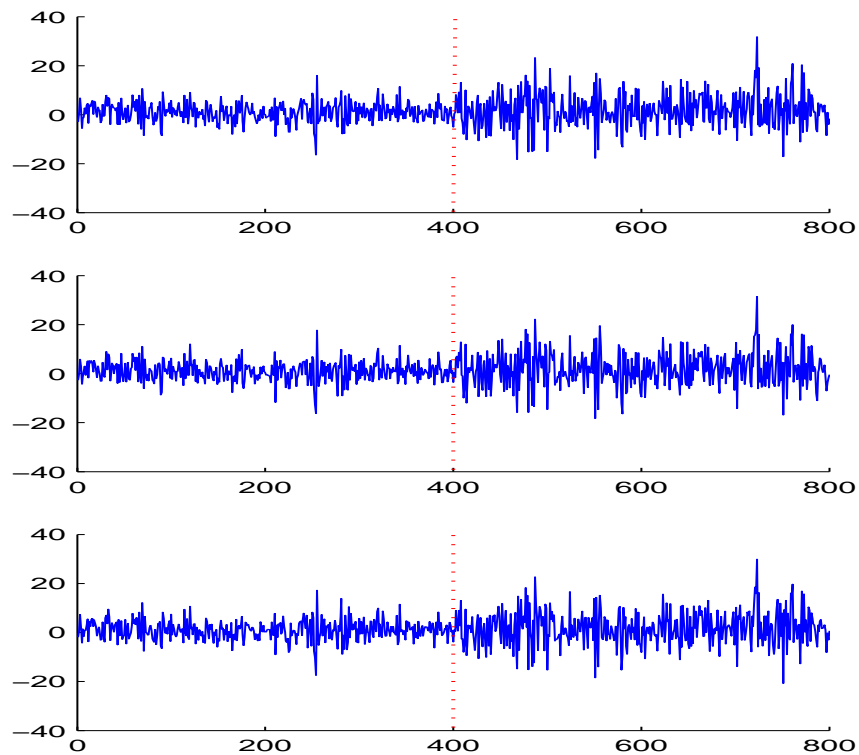
$$\log p(\mathcal{Z}|\mathcal{Y}, \hat{\Theta}_{\mathcal{M}}) = \max_{\mathcal{Z}} \left[\log p(\mathcal{Z}|\mathcal{Y}, \hat{\Theta}_{\mathcal{M}}) \right]$$

. EM Algorithm + GPB1, (.) EM Algorithm + Viterbi

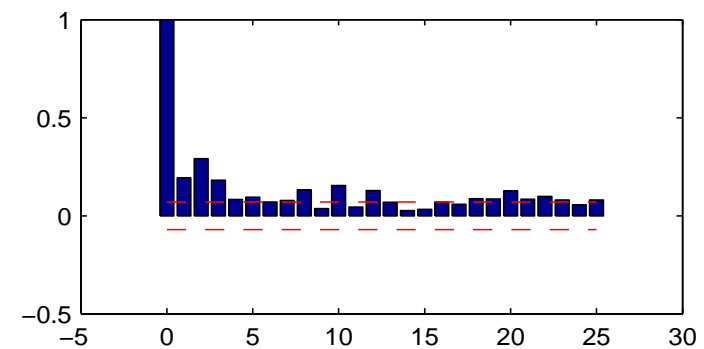
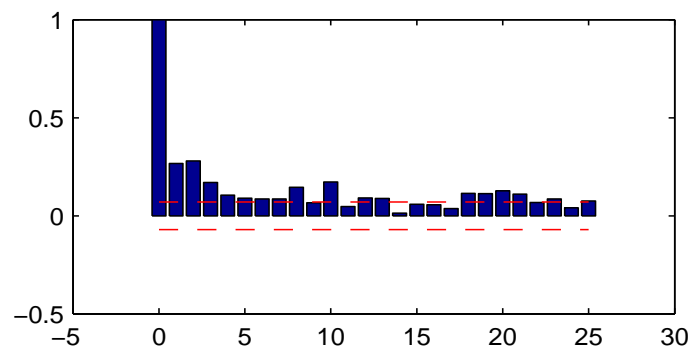
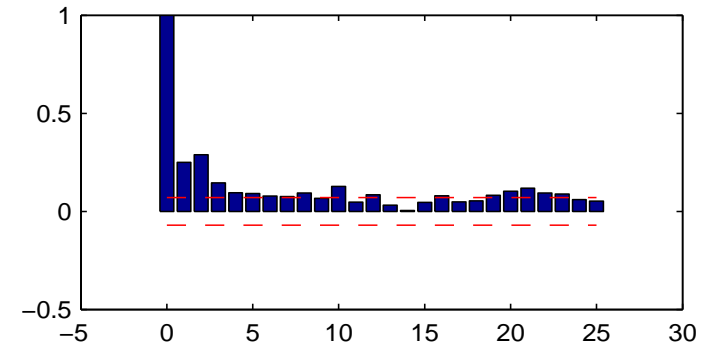
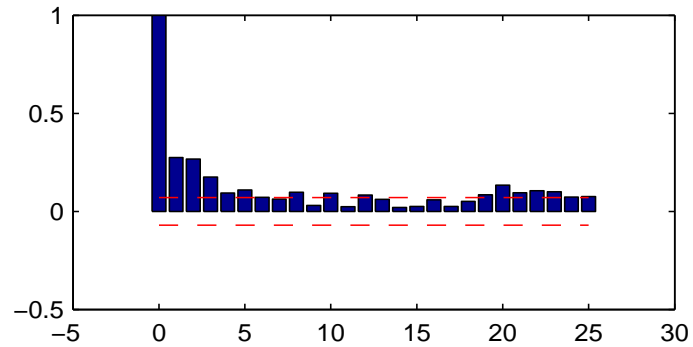
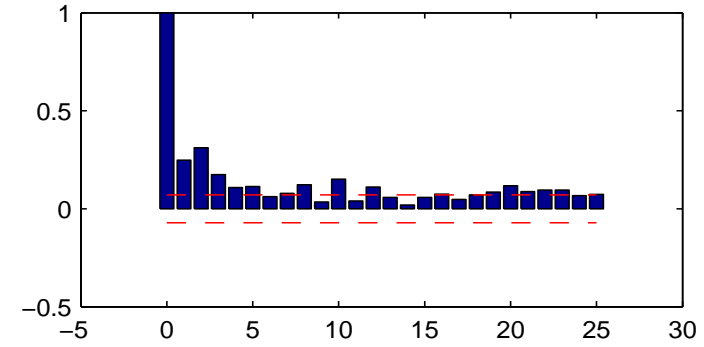
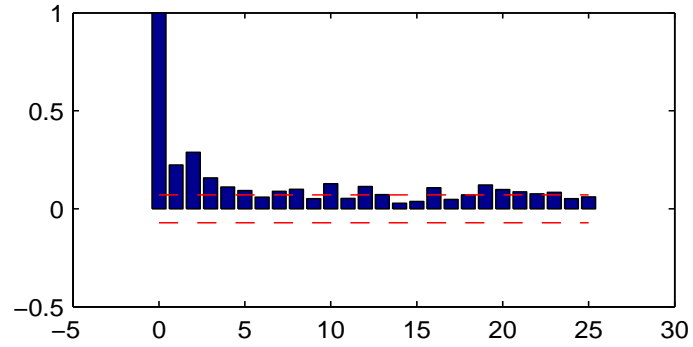
Critère	CHFAHMM	FAHMMNS	FAHMM1	FAHMM2	CHFA	FA1	FA2	FA3
AIC	90 (92)	10 (8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
BIC	96 (96)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
ICL	96 (96)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Sur la même base de données q'on a utilisé pour la génération des données d'un CHFAHMM, nous avons estimé des modèles FAHMM standards et non standards, un modèle CHFA et un modèle FA.

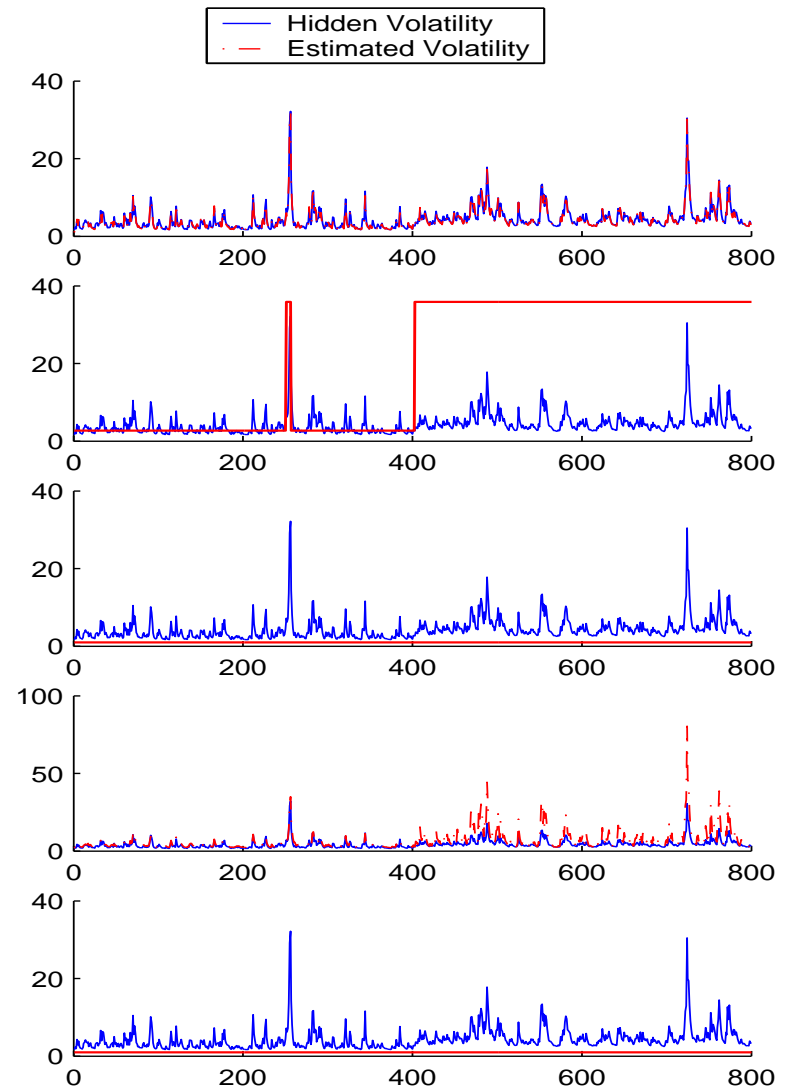
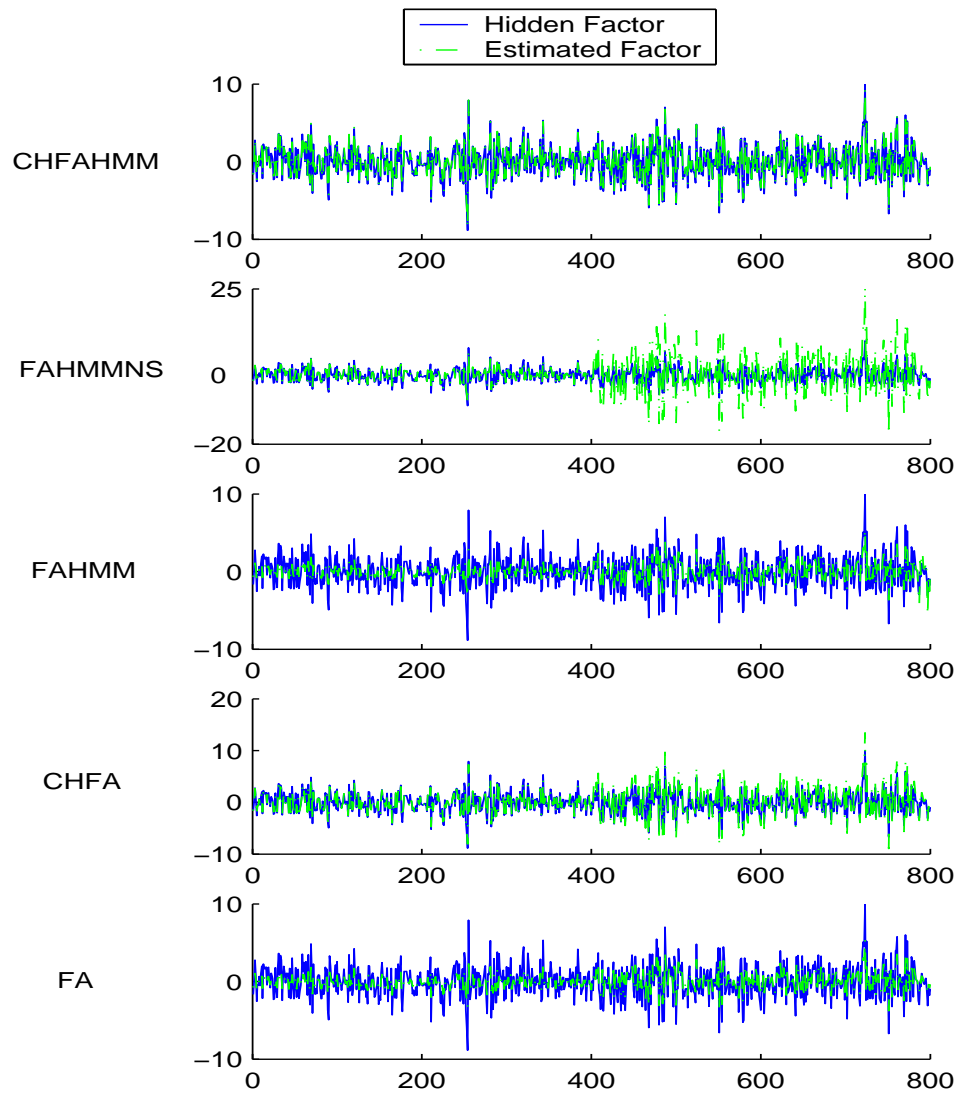
Les séries simulées



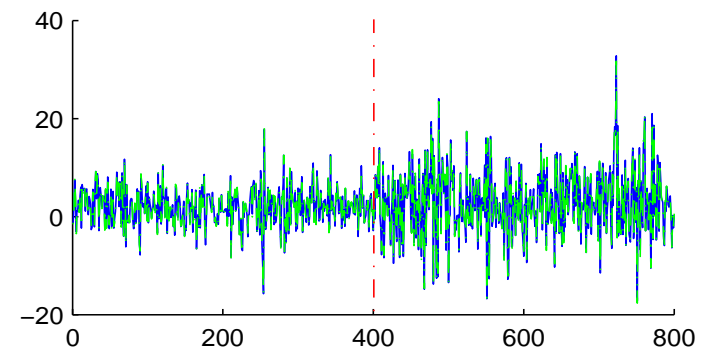
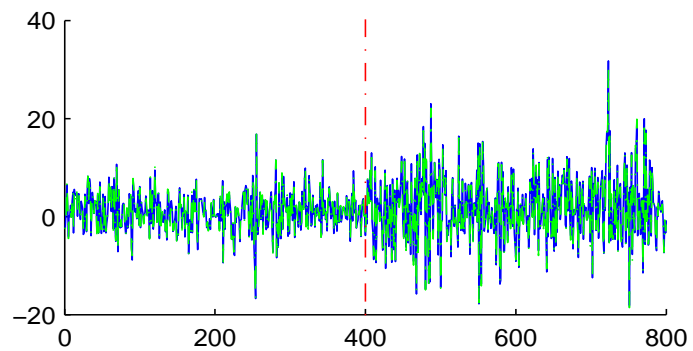
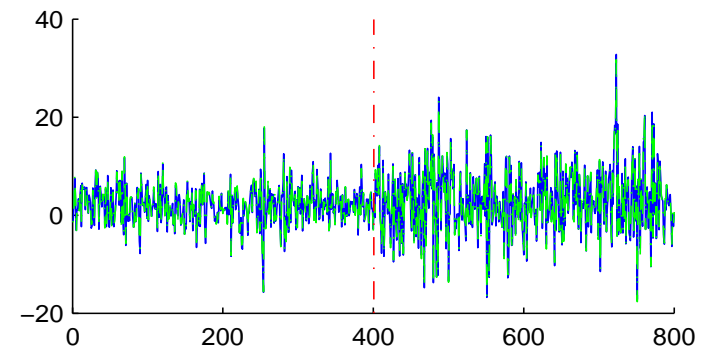
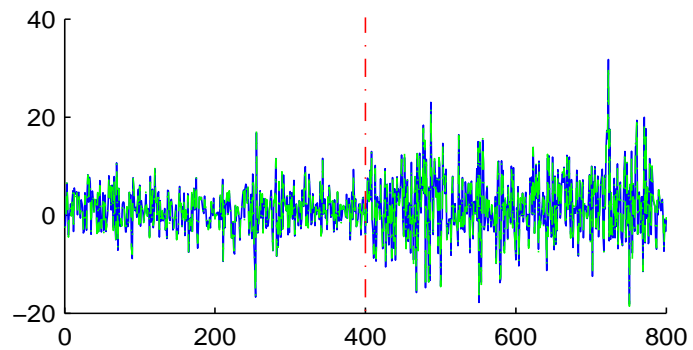
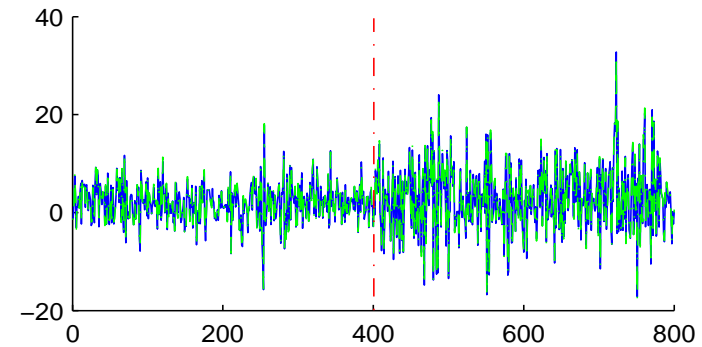
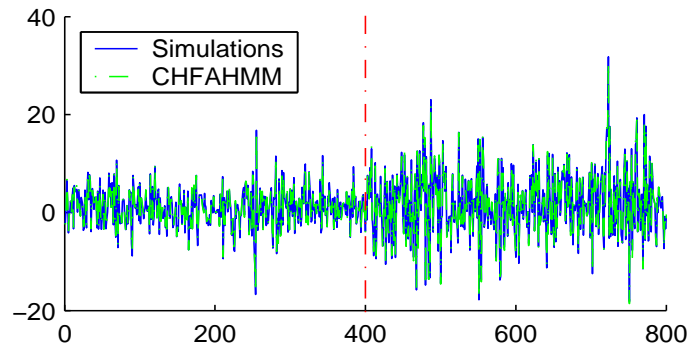
Fonction d'Autocorrélation des y_i^2



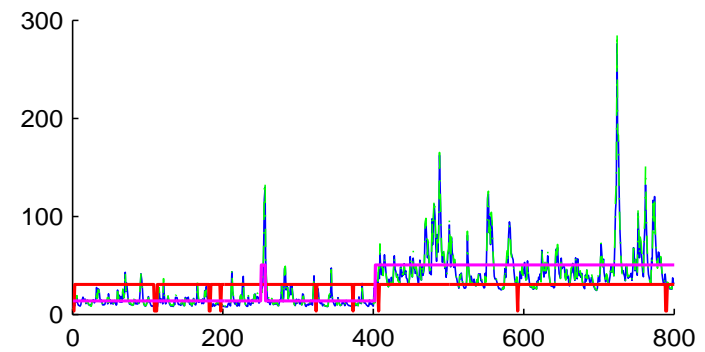
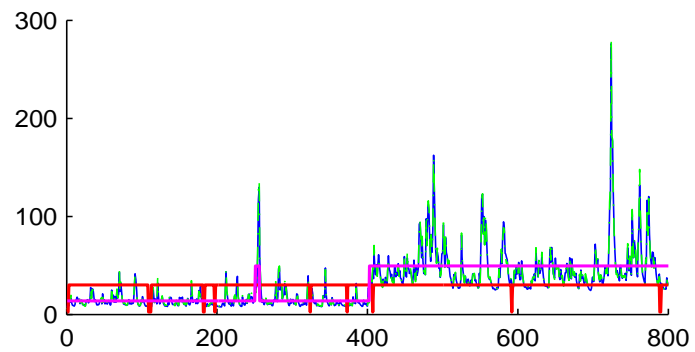
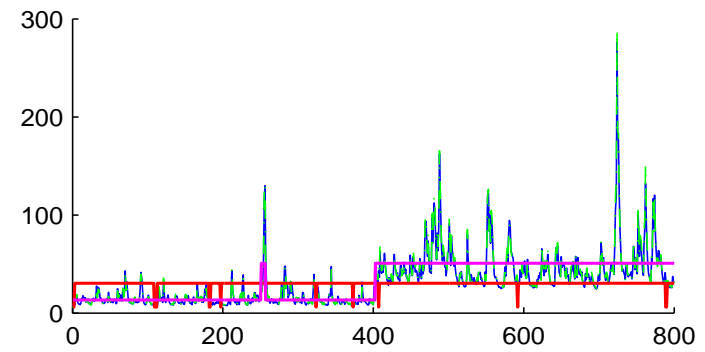
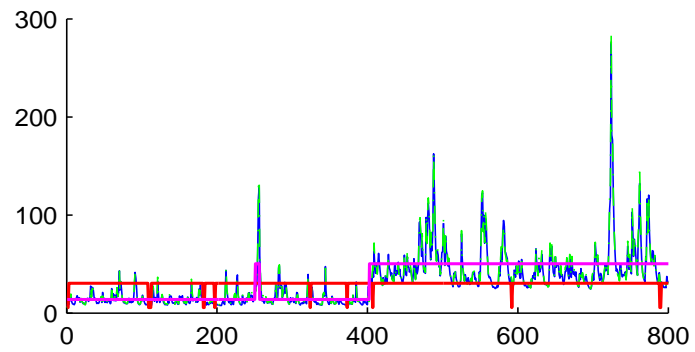
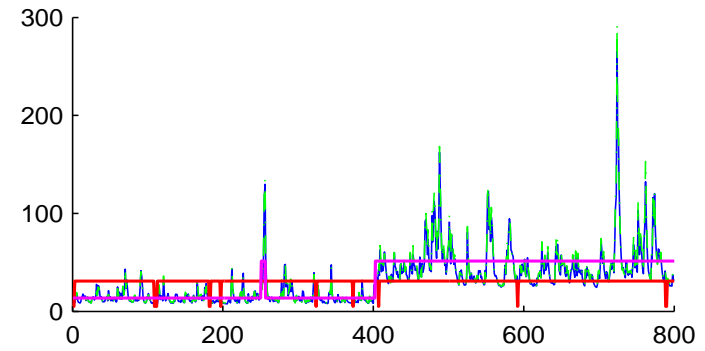
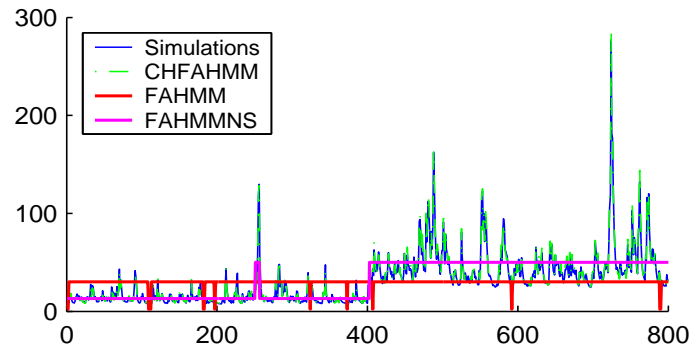
Volatilités et moyennes conditionnelles



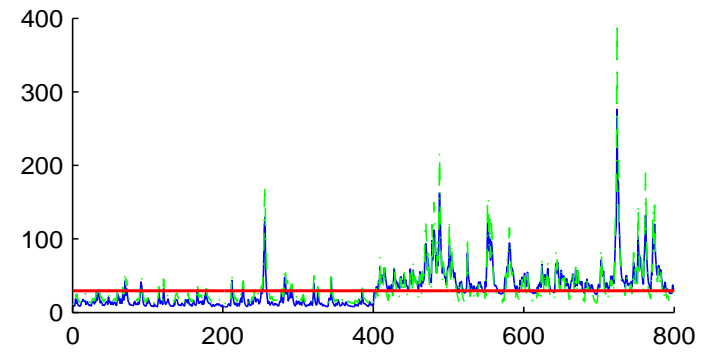
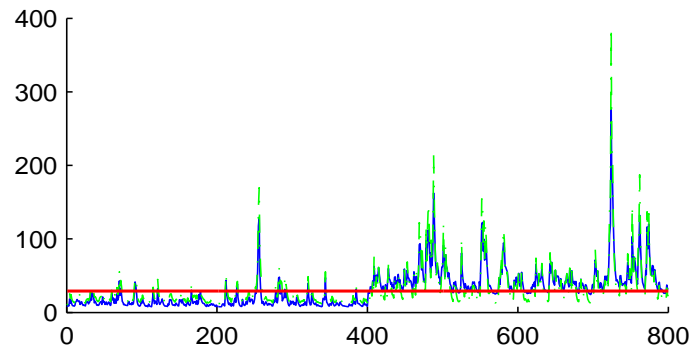
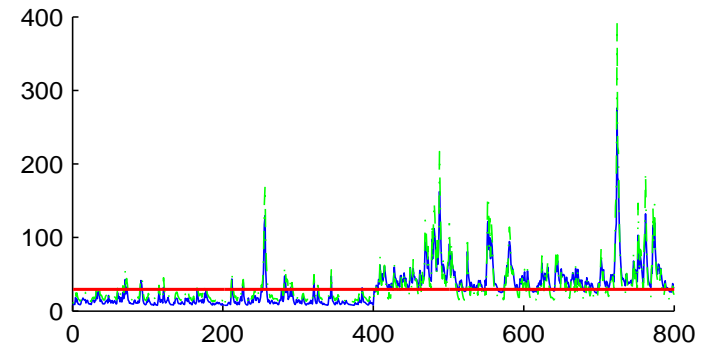
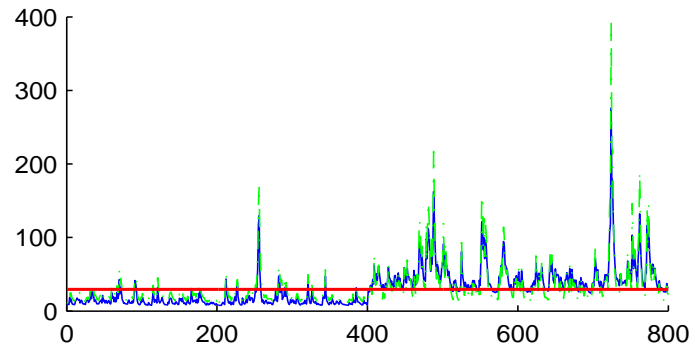
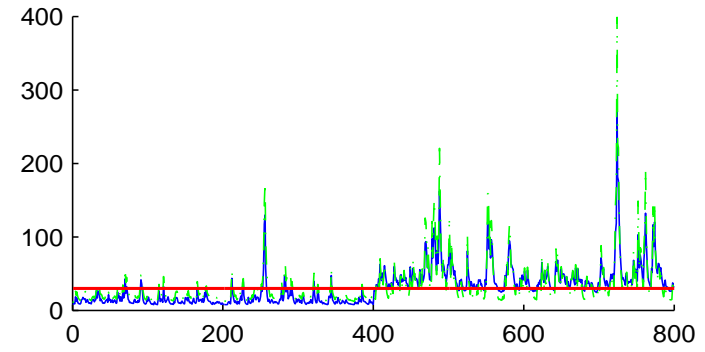
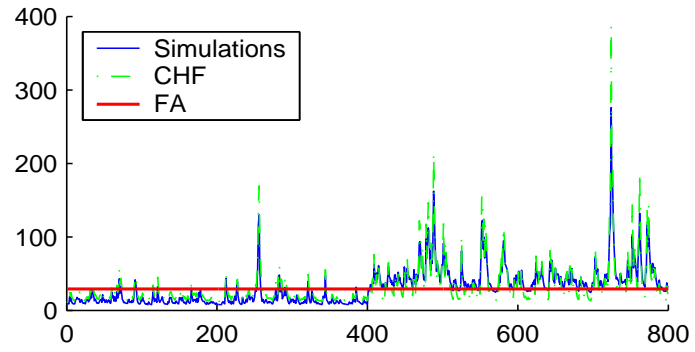
Moyennes conditionnelles des différentes séries : Modèle CHFAHMM



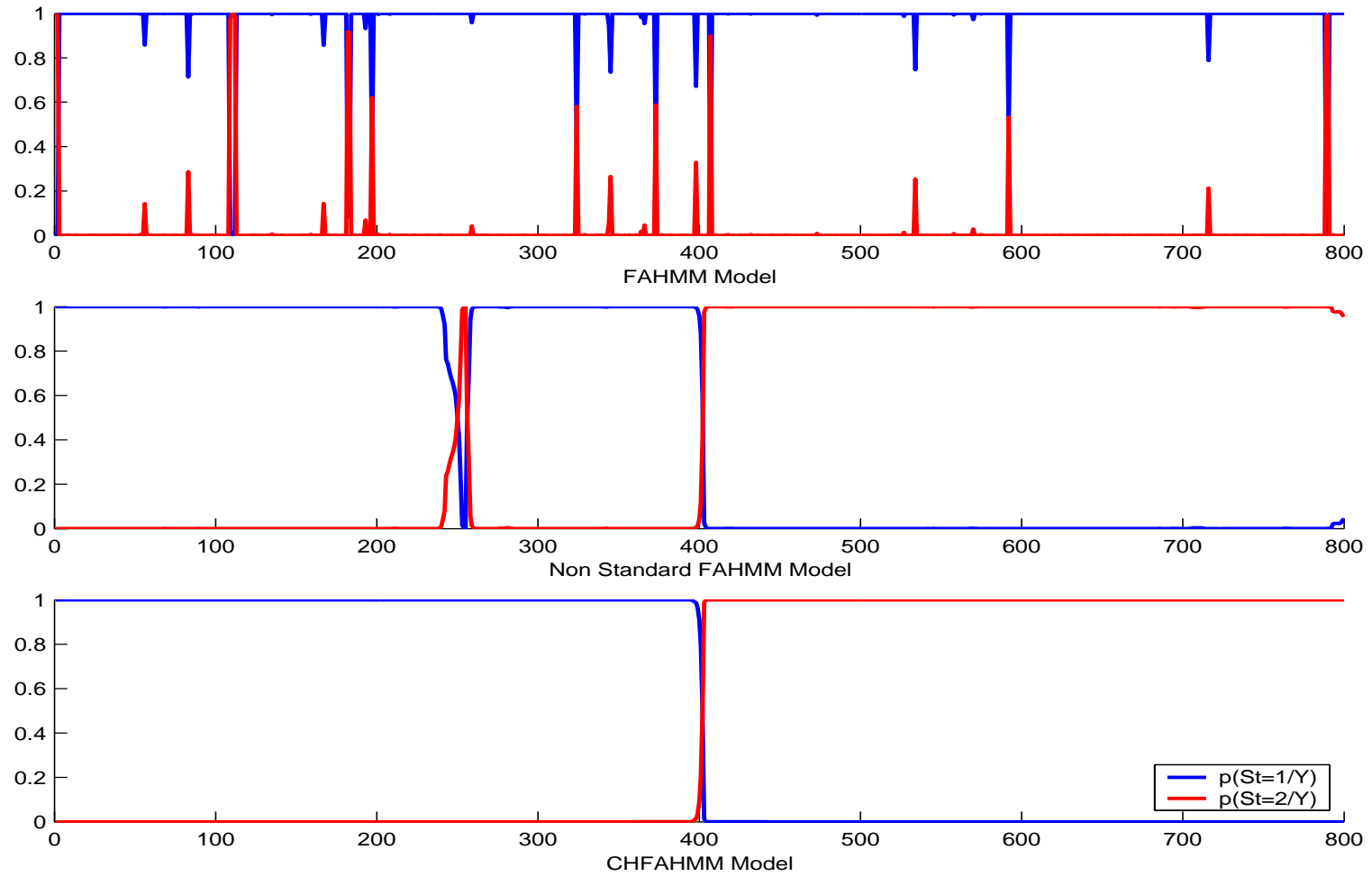
Volatilités conditionnelles des différentes séries



Volatilités conditionnelles des différentes séries

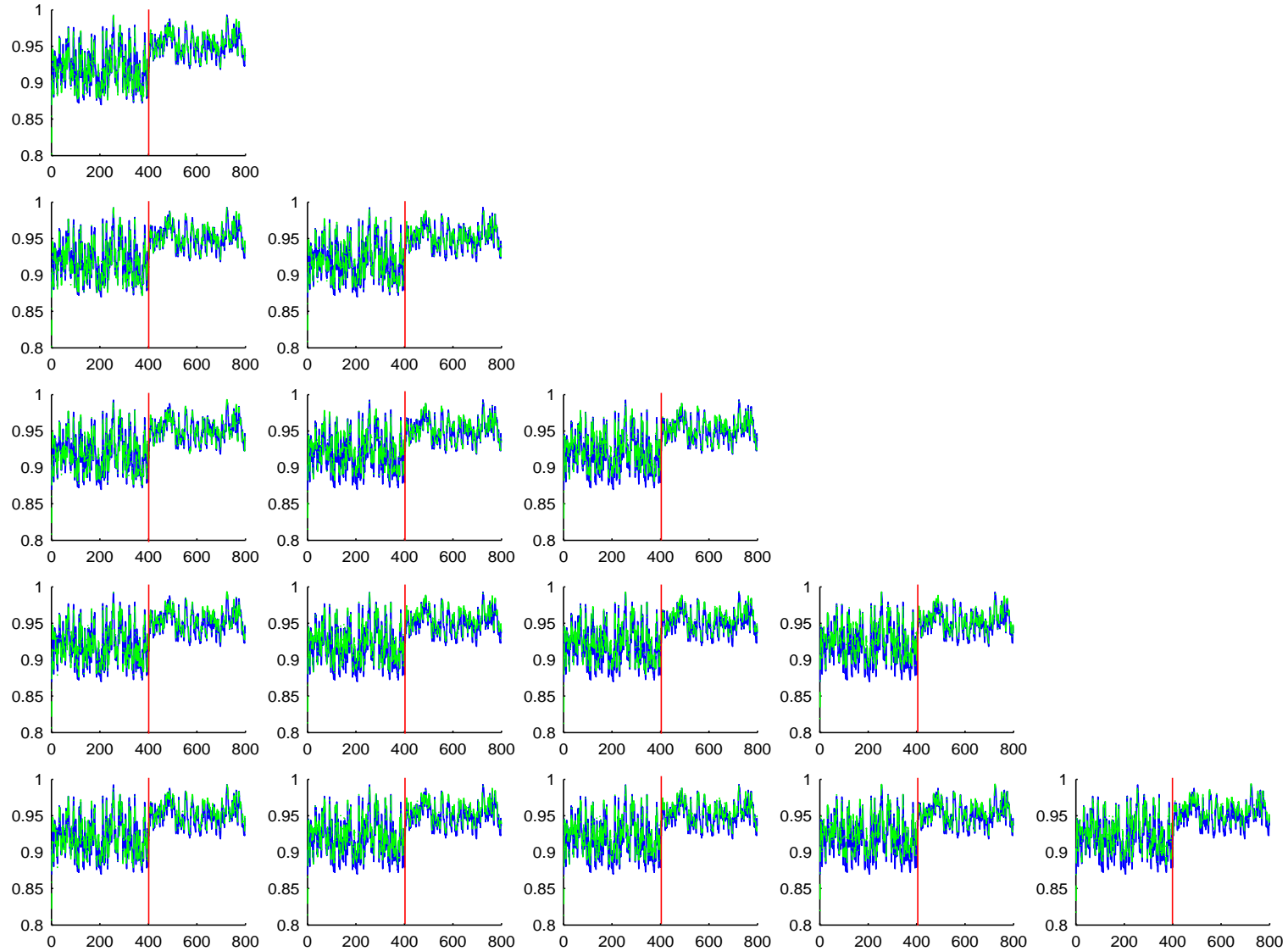


Probabilités a posteriori des états cachés



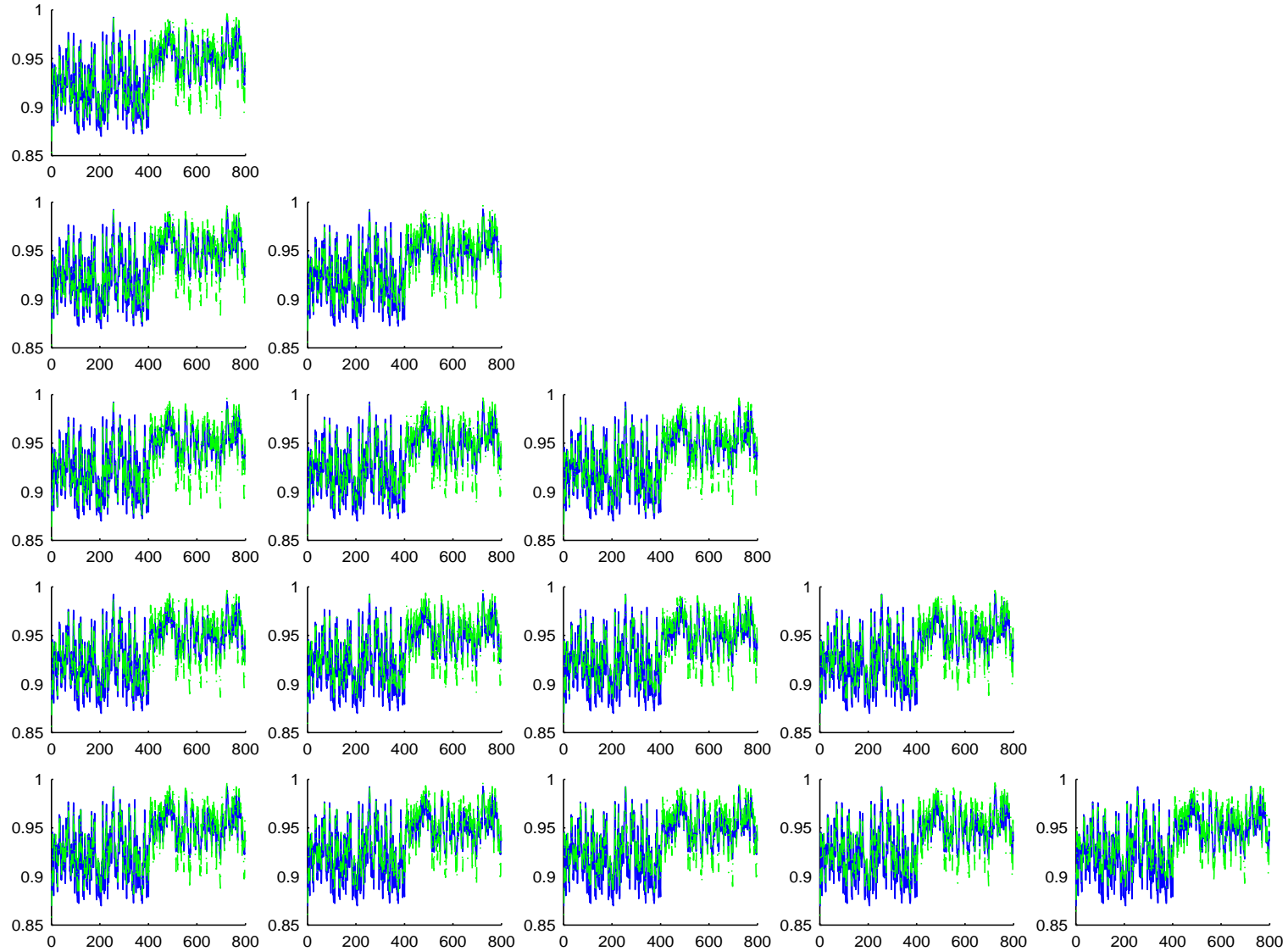
Structure de corrélation : Modèle CHFAHMM

✓ Simulations ✓ Estimations



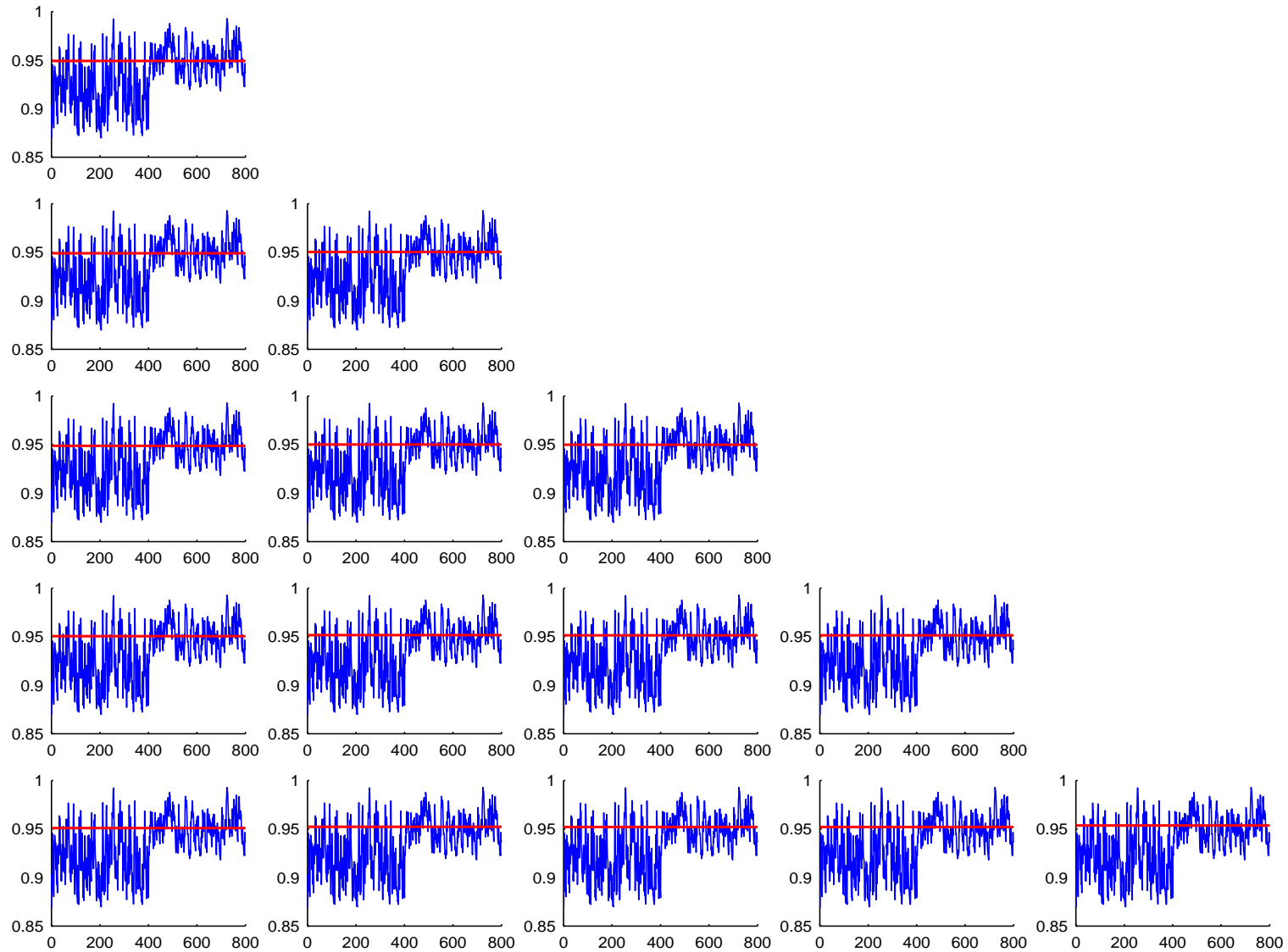
Structure de corrélation : Modèle CHFA

✓ Simulations ✓ Estimations



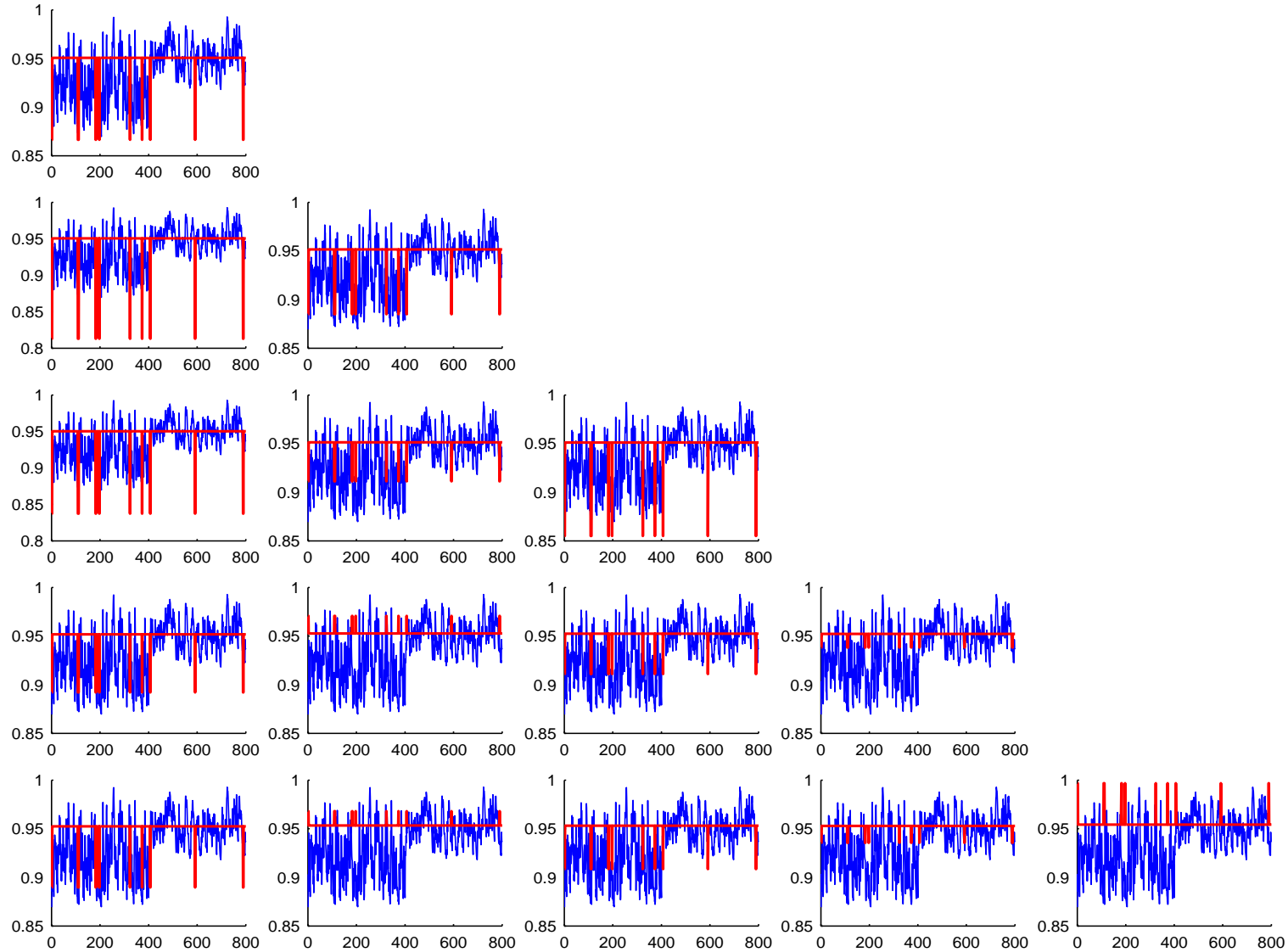
Structure de corrélation : Modèle FA

✓ Simulations ✓ Estimations



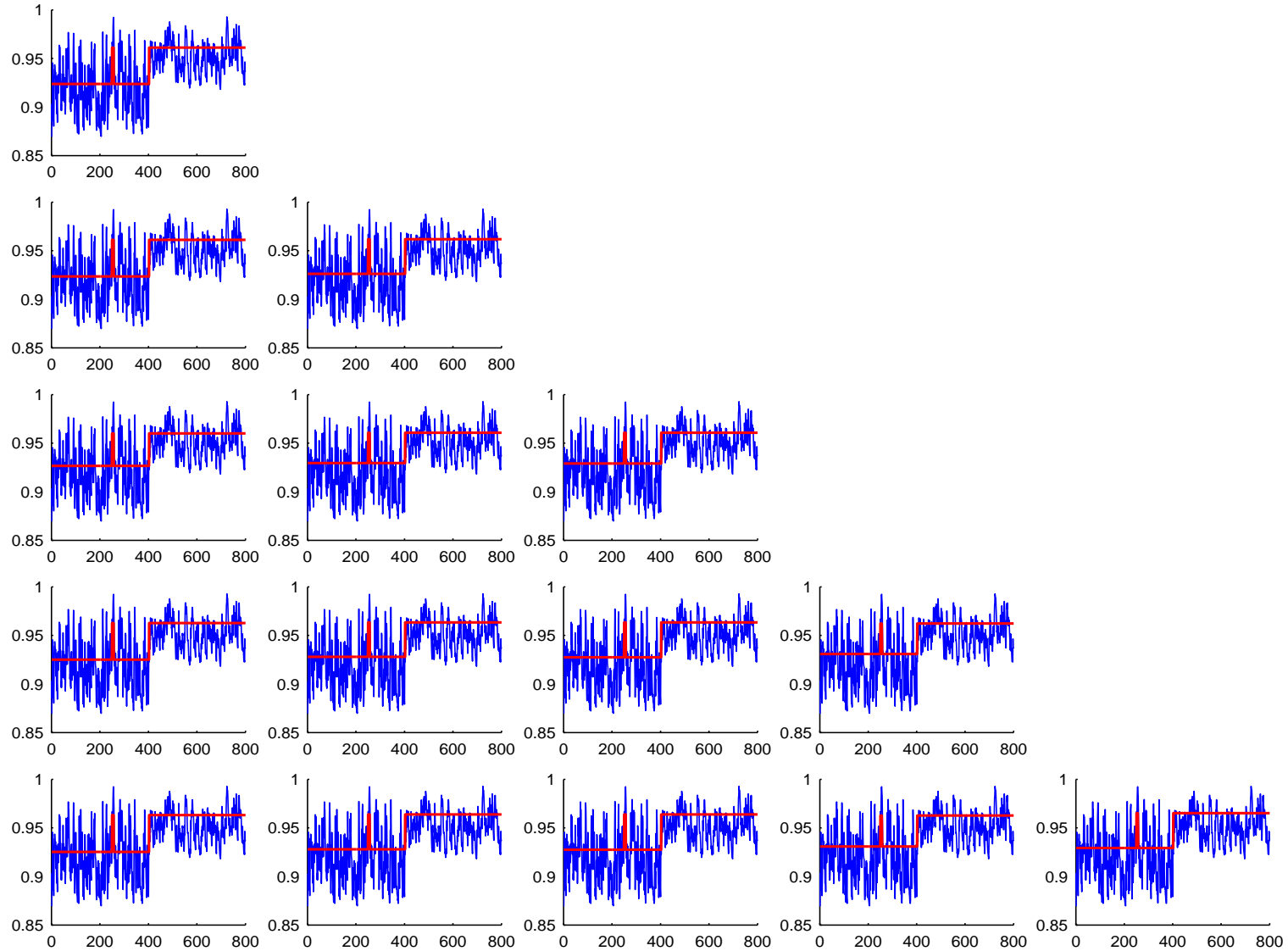
Structure de corrélation : Modèle FAHMM

✓ Simulations ✓ Estimations



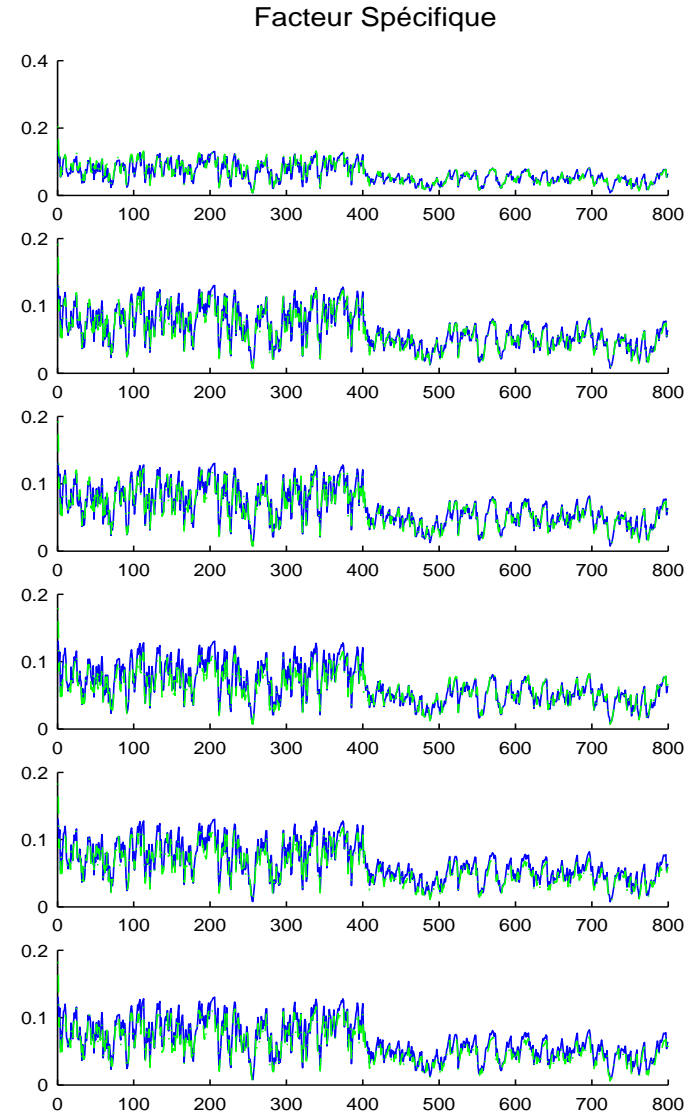
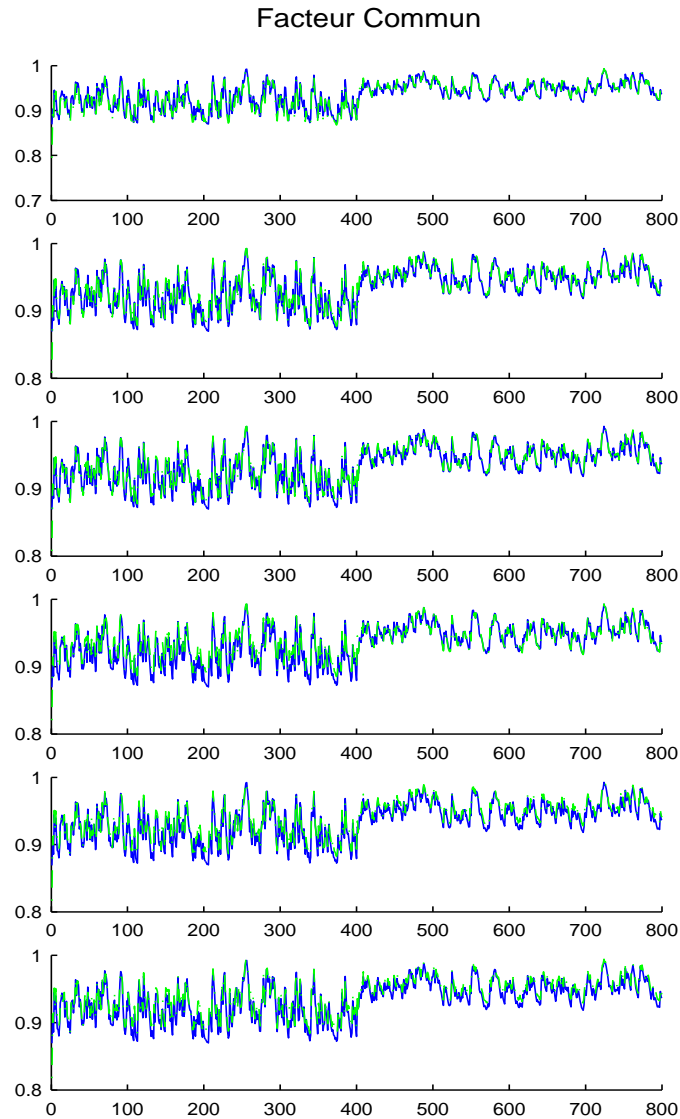
Structure de corrélation : Modèle FAHMMNS

✓ Simulations ✓ Estimations



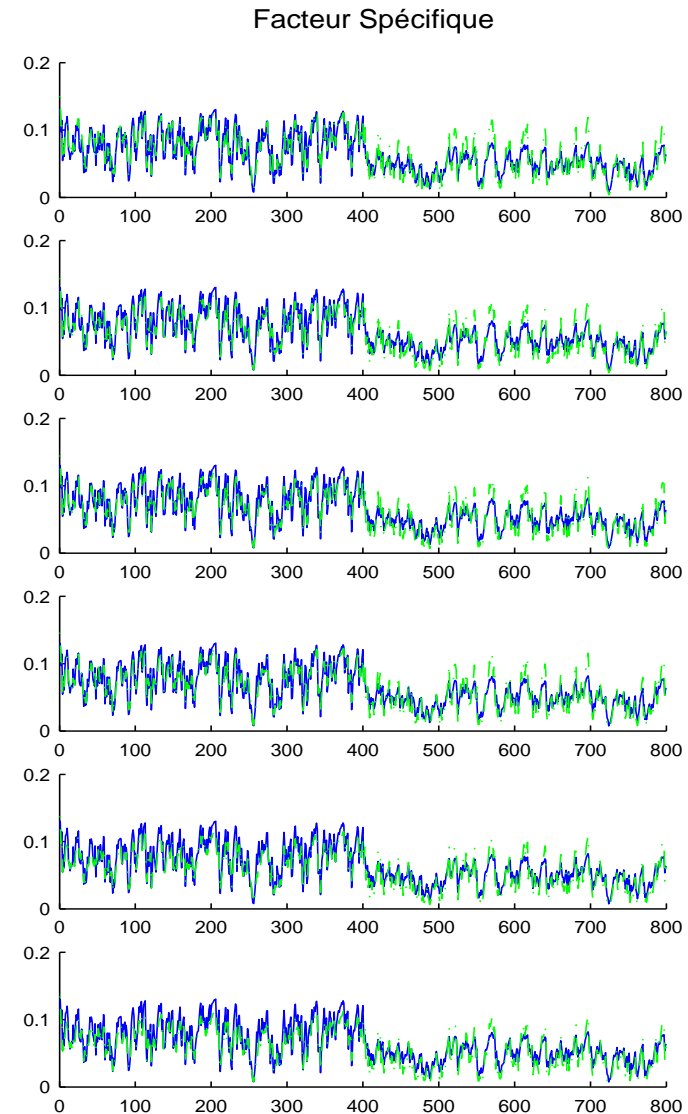
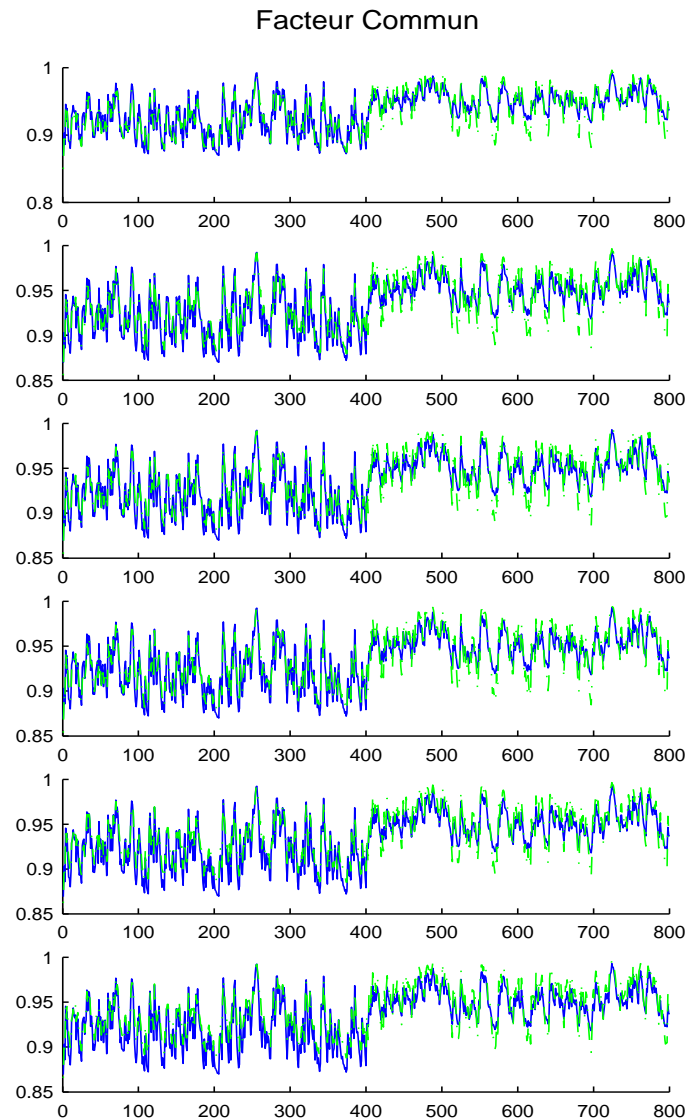
Part de la variance expliquée par chacun des facteurs : Modèle CHFAHMM

✓ Simulations ✓ Estimations



Part de la variance expliquée par chacun des facteurs : Modèle CHFA

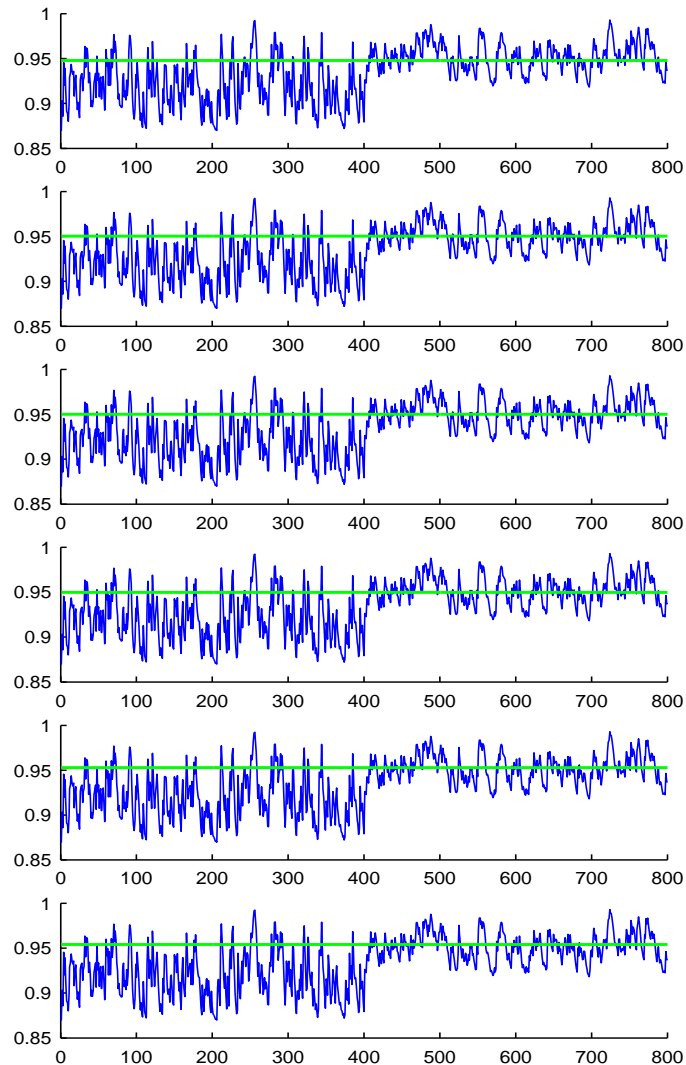
✓ Simulations ✓ Estimations



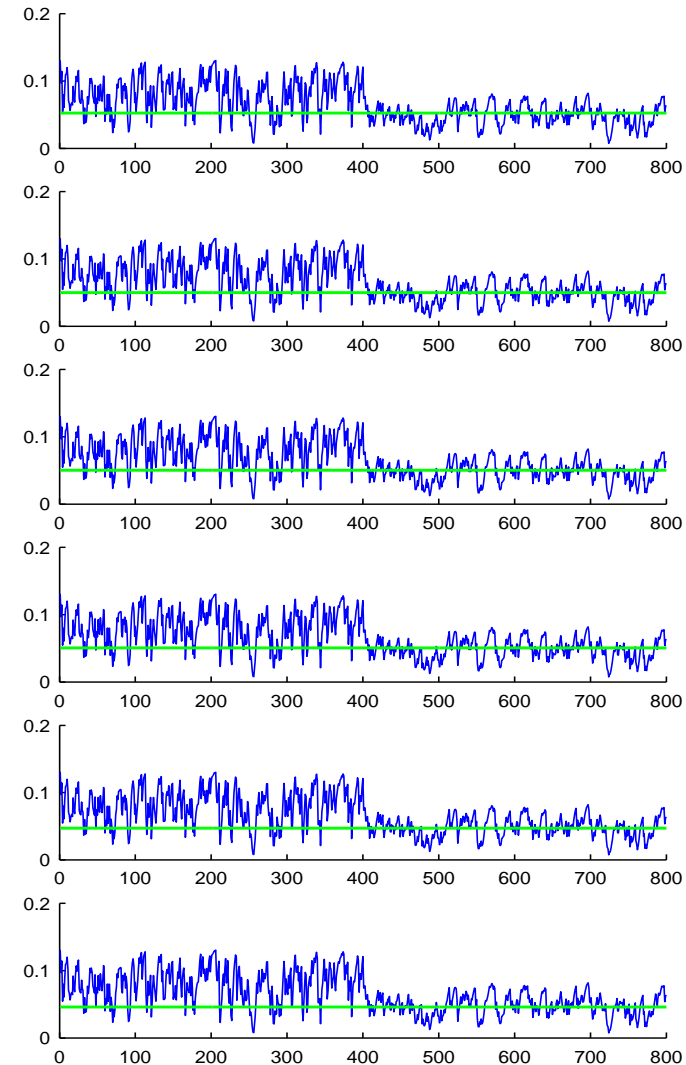
Part de la variance expliquée par chacun des facteurs : Modèle FA

✓ Simulations ✓ Estimations

Facteur Commun



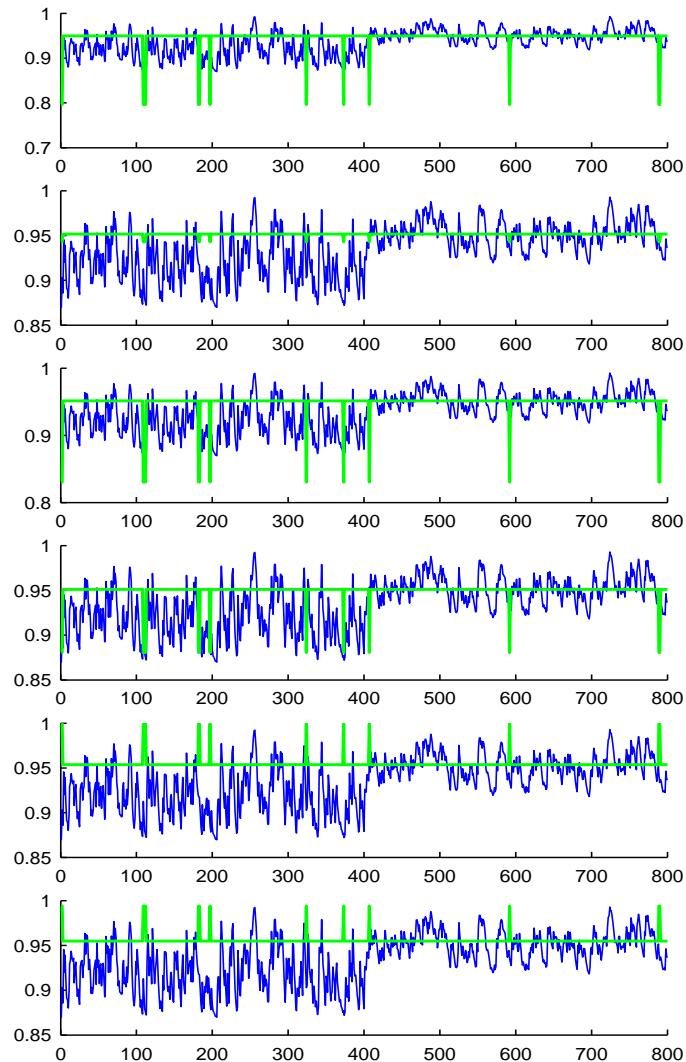
Facteur Spécifique



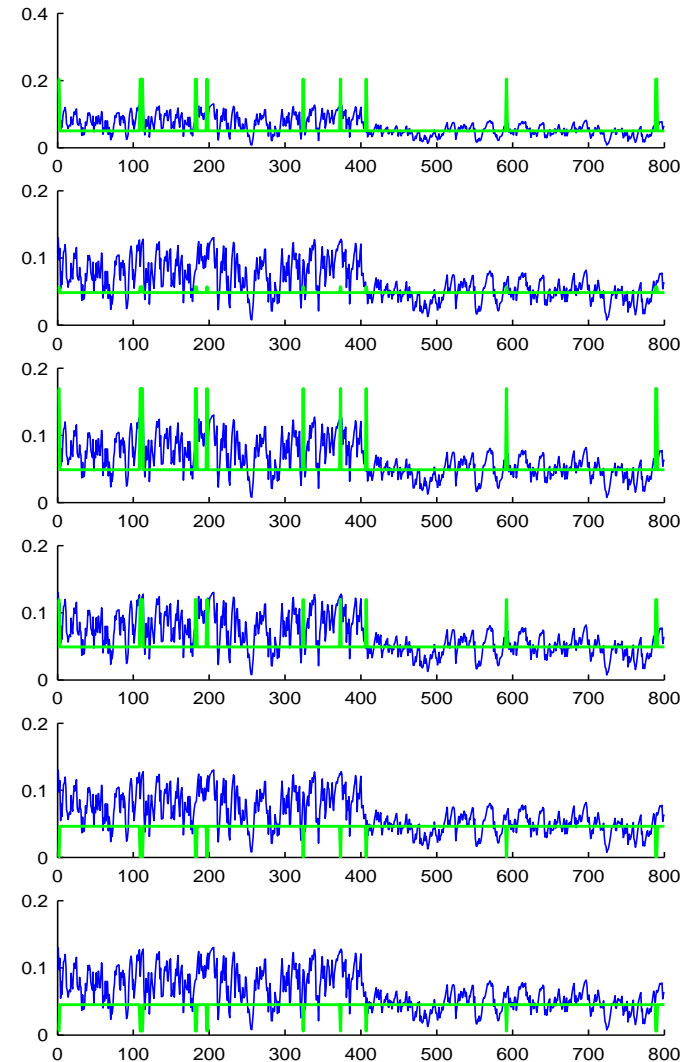
Part de la variance expliquée par chacun des facteurs : Modèle FAHMM

✓ Simulations ✓ Estimations

Facteur Commun



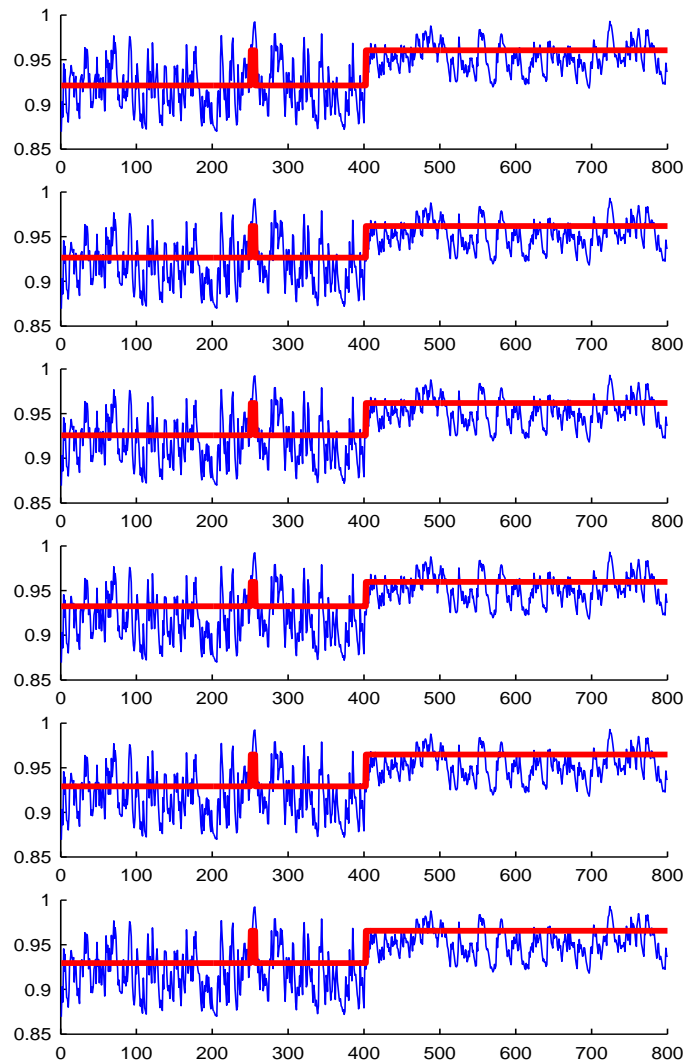
Facteur Spécifique



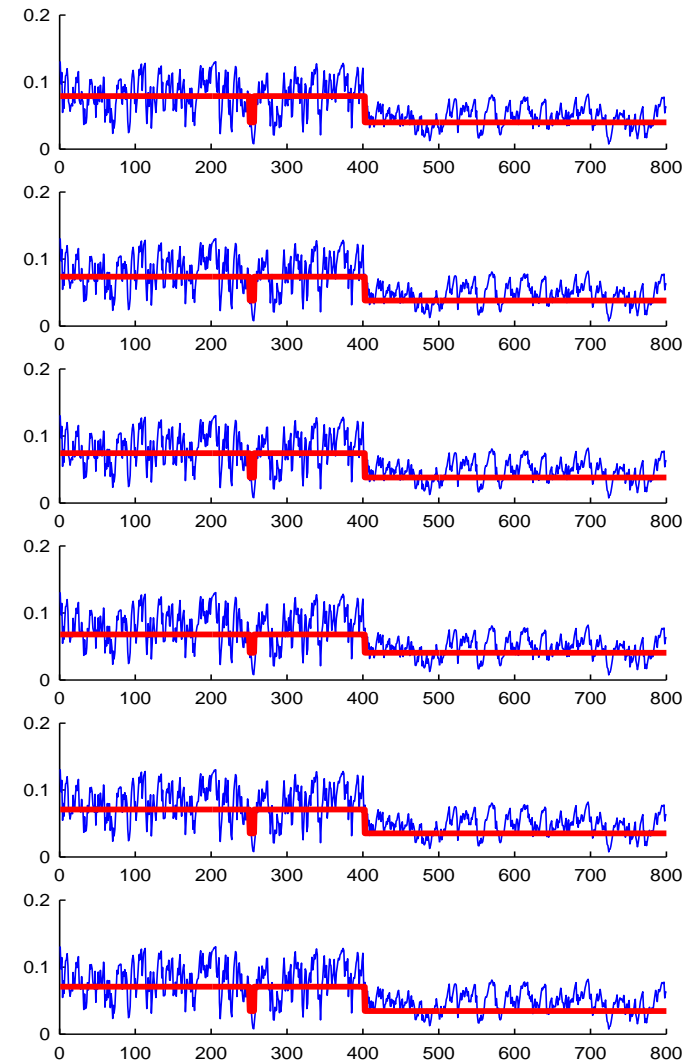
Part de la variance expliquée par chacun des facteurs : Modèle FAHMMNS

✓ Simulations ✓ Estimations

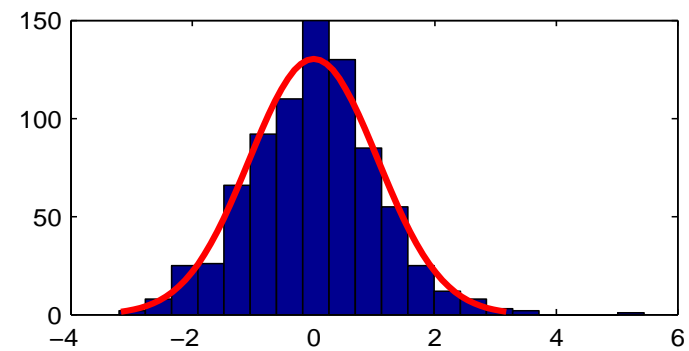
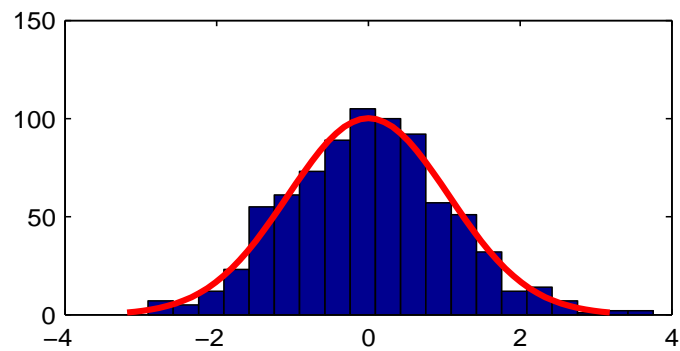
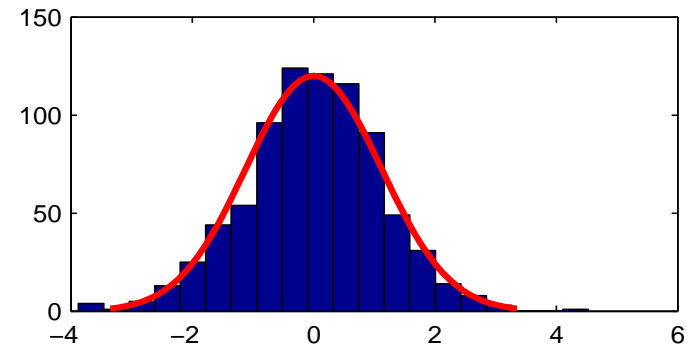
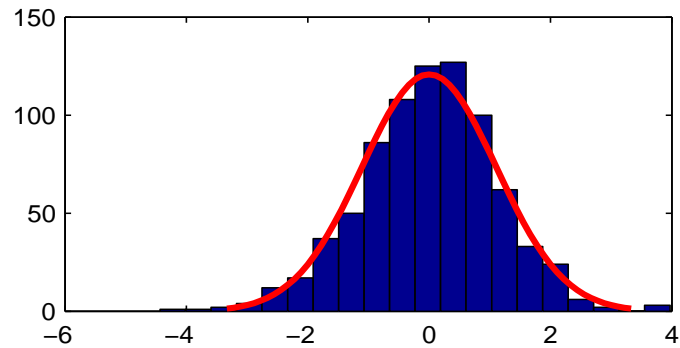
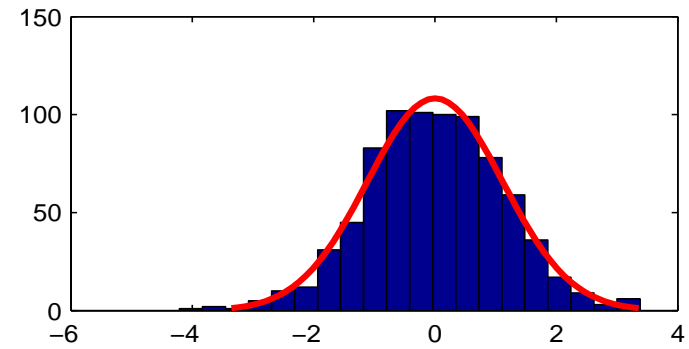
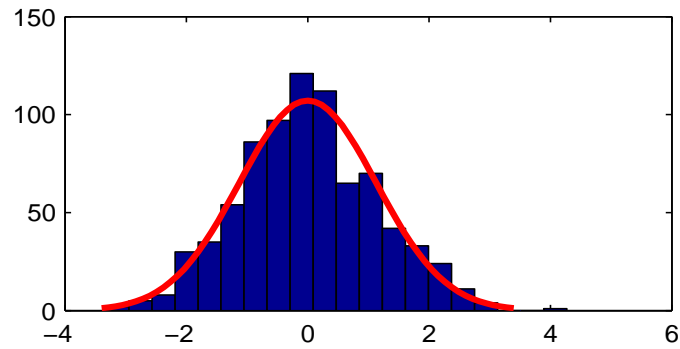
Facteur Commun



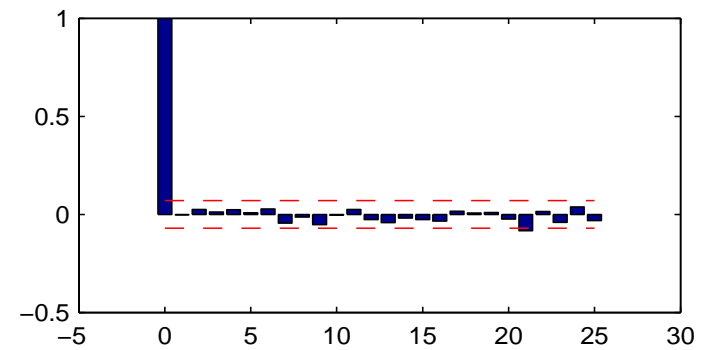
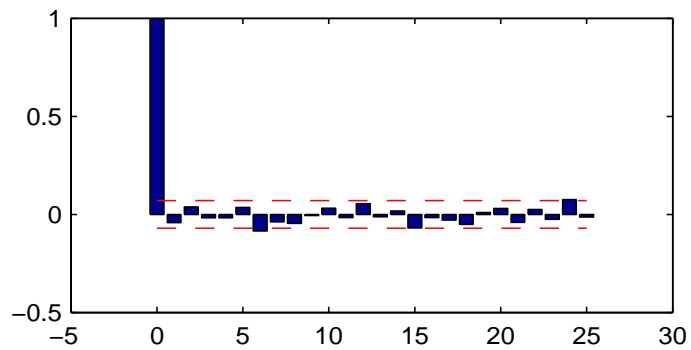
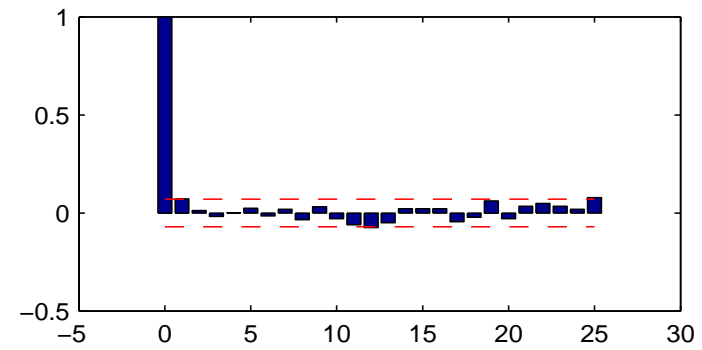
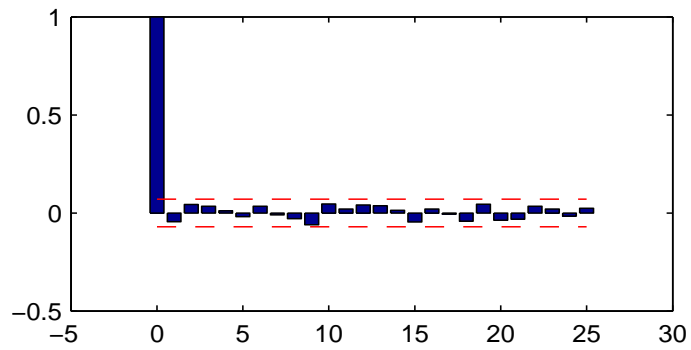
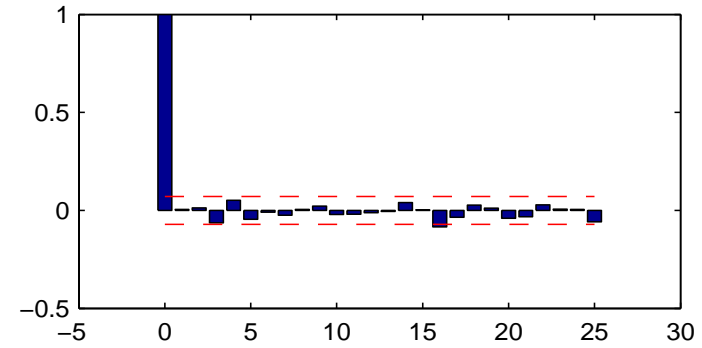
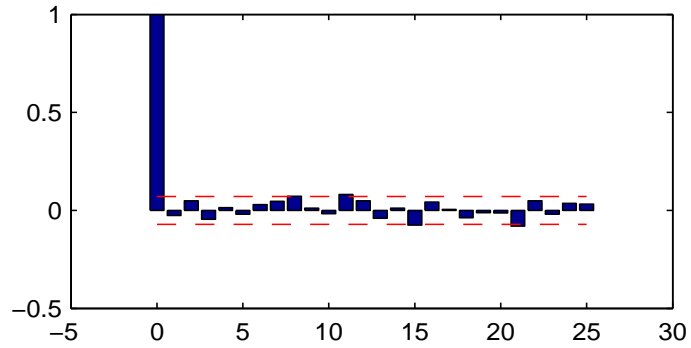
Facteur Spécifique



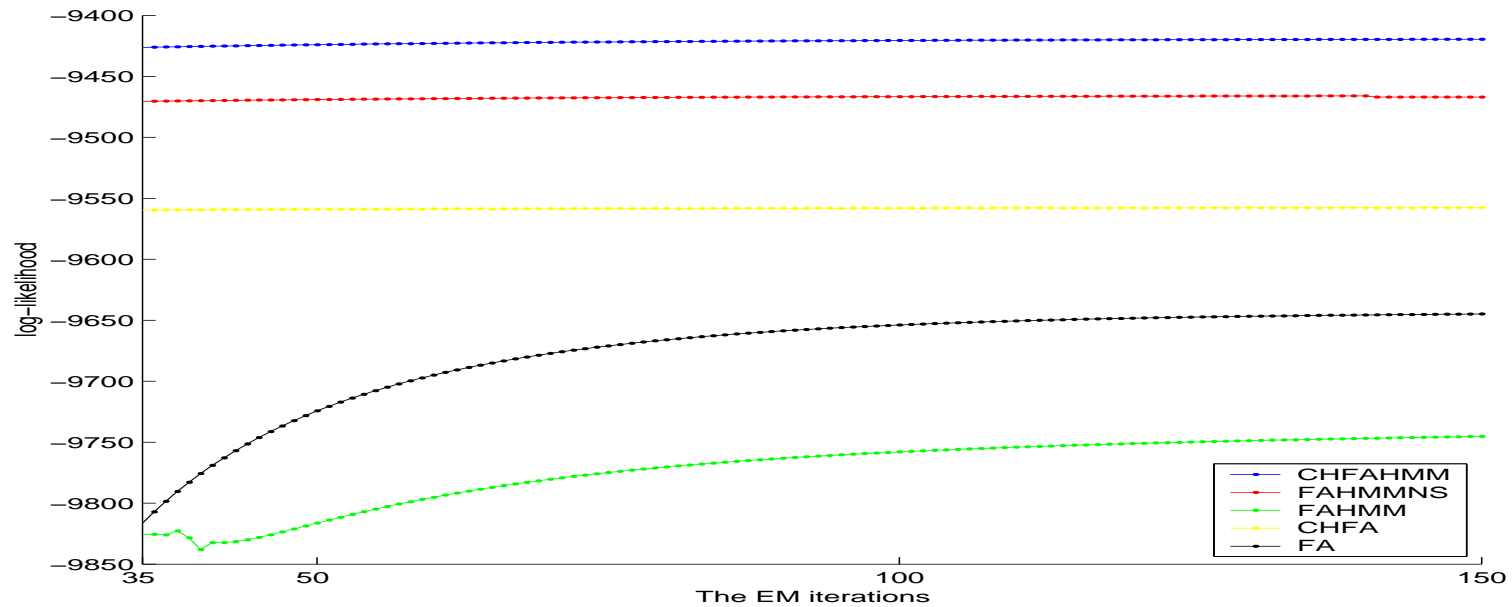
Distributions empiriques des erreurs : Modèle CHFAHMM



Fonctions d'autocorrélation des erreurs : Modèle CHFAHMM



log-vraisemblances des différents modèles



Critères de Sélection

Critères	CHFAHMM	FAHMMNS	FAHMM	CHFA	FA
AIC	18935	18982	19224	19157	19326
BIC	19160	19188	19421	19255	19410
ICL	19160	19189	19421	-	-