

# Estimation des paramètres et de l'état caché des modèles d'espace d'états: application au modèle d'Heston

El Kolei Salima sous la direction de F. Patras et P. Reynaud  
Bouret

Université de Nice Sophia Antipolis

CIRM 16-20 Avril 2012

$$\begin{cases} y_t = h_\theta(x_t, \varepsilon_t), & \text{Observable} \\ x_t = f_\theta(x_{t-1}, \eta_t) & \text{Inobservable} \end{cases}$$

- $(\varepsilon_t)$  et  $(\eta_t)$  sont des séquences de variables aléatoires indépendantes et identiquement distribuées.
- $h_\theta$  and  $f_\theta$  sont des fonctions dont la forme est connue et dépendent d'un vecteur de paramètre  $\theta \in \Theta$ .

## Hypothèses

- **(H1)** : Le signal  $\{x_t, t \in \mathbb{N}\}$  est modélisé par une *chaîne de Markov* de distribution initiale  $p_\theta(x_0)$  et de densité de transition  $p_\theta(x_t|x_{t-1})$ .
- **(H2)** : Les observations  $\{y_t, t \in \mathbb{N}^*\}$  sont supposées indépendantes conditionnellement au signal  $\{x_t, t \in \mathbb{N}\}$ , ie :  $p_\theta(y_t|y_{1:t-1}, x_{1:t}) = p_\theta(y_t|x_t)$ .

- Un modèle espace d'état est décrit par les quantités suivantes :
  - La distribution initiale :  $p_\theta(x_1)$ .
  - La densité de transition :  $p_\theta(x_t|x_{t-1})$ .
  - La vraisemblance :  $p_\theta(y_t|x_t)$ .
- **Supposons  $\theta$  connu : on veut estimer  $p_\theta(x_t|y_{1:t})$ ,  $\forall t \in [1, T]$  :**  
filtrage optimal  
 $\implies$  Le filtrage apparaît comme un outil “naturel” :  
mécanisme qui consiste à retirer l'information d'un processus non observable  $x_t$  (variable d'état) contenue dans des processus observables  $y_{1:t} = (y_1, \dots, y_t)$  (variables de mesure).

# Principe : "Bayes Recursion"

Le théorème de Bayes nous permet d'estimer récursivement la loi cible.

$$\underbrace{p_{\theta}(x_t|y_{1:t})}_{\text{distribution a posteriori}} = \frac{p_{\theta}(y_t|x_t)p_{\theta}(x_t|y_{1:t-1})}{p_{\theta}(y_t|y_{1:t-1})}$$
$$\propto p_{\theta}(y_t|x_t)p_{\theta}(x_t|y_{1:t-1}) \quad (1)$$

où

$$\underbrace{p_{\theta}(x_t|y_{1:t-1})}_{\text{distribution a priori}} = \int p_{\theta}(x_t|x_{t-1})p_{\theta}(x_{t-1}|y_{1:t-1})dx_{t-1} \quad (2)$$

Supposons :

- **(A1)** :  $f_\theta$  (resp.  $h_\theta$ ) linéaire en  $x_{t-1}$  et  $\eta_t$  (resp.  $x_t$  et  $\varepsilon_t$ ).
- **(A2)** :  $\varepsilon_t \sim \mathcal{N}(0, R)$  et  $\eta_t \sim \mathcal{N}(0, Q)$  avec  $R$  et  $Q$  connus.

Le modèle espace d'états peut s'écrire sous forme matricielle :

$$\begin{cases} y_t = C_\theta x_t + M_\theta \varepsilon_t & t \geq 1, \\ x_t = A_\theta x_{t-1} + G_\theta \eta_t, \end{cases}$$

$\implies$  On a une solution exacte pour l'estimation de  $p_\theta(x_t|y_{1:t})$  qui est donnée par le filtre de Kalman.

**Notation :**

- $\hat{x}_t^- = \mathbb{E}_\theta[x_t|y_{1:t-1}]$  et  $\hat{x}_t = \mathbb{E}_\theta[x_t|y_{1:t}]$ .
- $P_t^- = \text{Var}_\theta[(x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)'|y_{1:t-1}]$ .
- $P_t = \text{Var}_\theta[(x_t - \hat{x}_t)(x_t - \hat{x}_t)'|y_{1:t}]$ .

## Algorithme :

- **Étape de prédiction :**

$$\begin{cases} \hat{x}_t^- = A_\theta \hat{x}_{t-1} \\ P_t^- = A_\theta P_{t-1} A_\theta' + G_\theta Q G_\theta' \end{cases}$$

- **Étape de Mise à jour :**

$$\begin{cases} \hat{x}_t = \hat{x}_t^- + K_t [y_t - C_\theta \hat{x}_t^-] \\ P_t = (I - K_t C_\theta) P_t^- \\ K_t = P_t^- C_\theta' (C_\theta P_t^- C_\theta' + M_\theta R M_\theta')^{-1} \end{cases}$$

$K_t$  : matrice de Kalman : matrice qui minimise la variance a posteriori  $P_t$ .

# Modèle non linéaire et non Gaussien : Méthodes de Monte Carlo séquentielle pour le filtrage

- Méthode de Monte Carlo basée sur une simulation importante de points notés  $(x_t^i)_{i=1, \dots, M}$  pondérés par des poids  $w_t^i$ . “ $M$ ” représente le nombre points simulés.
- Un filtre particulaire est une collection de ces points pondérés appelés particules.

## Contexte

- Supposons que nous sommes intéressé par  $p_\theta(x_t | y_{1:t})$ .
- On veut obtenir  $(w_t^i, x_t^i)$  qui suit la loi  $p_\theta(x_t | y_{1:t})$ .

- **Distribution d'importance** :  $q_\theta(x_t|x_{1:t-1}, y_{1:t})$
- Si  $x_t^i$  est de loi  $q_\theta(x_t|x_{1:t-1}, y_{1:t})$  alors  $\sum_{i=1}^M w_{t-1}^i \delta_{x_t^i}$  converge vers la loi  $q_\theta(x_t|x_{1:t-1}, y_{1:t})p_\theta(x_{t-1}|y_{1:t-1})$ .
- On calcule les poids  
 $w_t^i \propto w_{t-1}^i p_\theta(y_t|x_t)p_\theta(x_t|x_{t-1})/q_\theta(x_t|x_{1:t-1}, y_{1:t})$ .
- Finalement on a  $\sum_{i=1}^M w_t^i \delta_{x_t^i}$  converge vers  $p_\theta(x_t|y_{1:t})$  (Del Moral, Doucet...).
- Pour éviter le problème de dégénérescence des poids : on ajoute une étape de rééchantillonnage.

## Algorithme :

- **Pour**  $t = 1$  **et**  $i = 1 \dots M$  :

- Simuler  $x_1^i \sim q_\theta(x_1)$  et calculer  $w_1^i = \frac{p_\theta(x_1)}{q_\theta(x_1)}$ .
- Normaliser les poids  $\tilde{w}_1^i = \frac{w_\theta(x_1^i)}{\sum_{j=1}^M w_1^j}$ .

- **Pour**  $t = 2 \dots T$  **et**  $i = 1 \dots M$  :

- 1 Simuler  $x_t^i \sim q_\theta(x_t | x_{1:t-1}^i, y_{1:t})$  et poser  $x_{1:t}^i = (x_{1:t-1}^i, x_t^i)$ .
- 2 Mise à jour :

$$w_t^i = \frac{p_\theta(y_t | x_t^i) p_\theta(x_t^i | x_{t-1}^i)}{q_\theta(x_t^i | x_{1:t-1}^i, y_{1:t})}$$

- 3 Normaliser les poids :  $\tilde{w}_t^i$ .
- 4 Rééchantillonnage : On tire :  $\{x_t^i, \tilde{w}_t^i\}$ .
- 5 Poser  $\tilde{w}_t^i = 1/M$ .

En pratique, on choisit  $q_\theta(x_t | x_{1:t-1}^i, y_{1:t}) = p_\theta(x_t | x_{t-1}^i)$ .

## Algorithme :

- **Pour**  $t = 1$  **et**  $i = 1 \dots M$  :
  - Simuler  $x_1^i \sim p_\theta(x_1)$  et calculer  $\tilde{w}_1^i = \frac{1}{M}$ .
- **Pour**  $t = 1 \dots n$  **et**  $i = 1 \dots M$  :
  - 1 Calculer  $\mu_t^i = \mathbb{E}_\theta(x_t | x_{t-1}^i)$  et poser  $w_{t-1}^i = \tilde{w}_{t-1}^i p_\theta(y_t | \mu_t^i)$ .
  - 2 Rééchantillonner :
    - On rééchantillonne  $\{x_{t-1}^i, w_{t-1}^i\}$
    - On obtient  $\{\tilde{x}_{t-1}^i, 1/M\}$  et un échantillon d'indices  $\{j^i\}$  (correspond aux particules rééchantillonnées).
  - 3 Propager :
    - Poser  $x_{t-1}^i = \tilde{x}_{t-1}^i$ .
    - Simuler  $x_t^i \sim p_\theta(x_t | x_{t-1}^i)$ .
    - Calculer  $w_t^i = \frac{p_\theta(y_t | x_t^i)}{p_\theta(y_t | \mu_t^{j^i})}$  et normaliser  $\tilde{w}_t^i$ .

**On veut estimer simultanément et séquentiellement  $\theta$  et  $x_t$  au vu des observations.**

- Approche de Kitagawa, augmenter le vecteur d'état :

$$\mathbf{x}_t = (x_t, \theta_t), \quad \theta_t = \theta_{t-1}, \quad \theta_0 \sim p(\theta_0)$$

- ...couplée avec l'idée de Gordon :

$$\theta_t = \theta_{t-1} + \mathcal{G}_t, \quad \mathcal{G}_t \sim \mathcal{N}(0, \mathcal{W}_t)$$

et  $\theta_{t-1}$  indépendant de  $\mathcal{G}_t$ .

**Avantage :** La méthode marche mieux en pratique que de considérer  $\mathbf{x}_t = (x_t, \theta_t)$  avec  $\theta_t = \theta_{t-1}$ .

Paramètre dynamique :  $\theta_t = \theta_{t-1} + \mathcal{N}(0, \mathcal{W}_t)$

① Prediction :

- Calculer  $\mu_t^i = \mathbb{E}(x_t | x_{t-1}^i, \theta_{t-1}^i)$  et  $m_{t-1}^i = \mathbb{E}[\theta_t | \theta_{t-1}^i] = \theta_{t-1}^i$
- Poser  $w_{t-1}^i = \tilde{w}_{t-1}^i p(y_t | \mu_t^i, m_{t-1}^i)$ .

② Rééchantillonnage :

③ Propager :

- Poser  $x_{t-1}^i = \tilde{x}_{t-1}^i$  et  $\theta_{t-1}^i = \theta_{t-1}^i$ .
- Simuler  $\theta_t^i \sim \mathcal{N}(m_{t-1}^i, \mathcal{W}_t)$
- Simuler  $x_t^i \sim p(x_t | x_{t-1}^i, \theta_t^i)$ .
- Calculer les poids :  $w_t = p(y_t | x_t^i, \theta_t^i) / p(y_t | \mu_t^i, m_{t-1}^i)$  et normaliser  $\tilde{w}_t$ .

Paramètre dynamique :  $\theta_t = \theta_{t-1} + \mathcal{N}(0, \mathcal{W}_t)$

① Prediction :

- Calculer  $\mu_t^i = \mathbb{E}(x_t | x_{t-1}^i, \theta_{t-1}^i)$  et  $m_{t-1}^i = \mathbb{E}[\theta_t | \theta_{t-1}^i] = \theta_{t-1}^i$
- Poser  $w_{t-1}^i = \tilde{w}_{t-1}^i p(y_t | \mu_t^i, m_{t-1}^i)$ .

② Rééchantillonnage :

③ Propager :

- Poser  $x_{t-1}^i = \tilde{x}_{t-1}^i$  et  $\theta_{t-1}^i = \theta_{t-1}^i$ .
  - Simuler  $\theta_t^i \sim \mathcal{N}(m_{t-1}^i, \mathcal{W}_t)$
  - Simuler  $x_t^i \sim p(x_t | x_{t-1}^i, \theta_t^i)$ .
  - Calculer les poids :  $w_t = p(y_t | x_t^i, \theta_t^i) / p(y_t | \mu_t^i, m_{t-1}^i)$  et normaliser  $\tilde{w}_t$ .
- 
- **Problème** :  $\text{Var}(\theta_t) = \text{Var}(\theta_{t-1}) + \mathcal{W}_t$  : Les variances s'accroissent ! Perte d'information !

Paramètre dynamique :  $\theta_t = \theta_{t-1} + \mathcal{N}(0, \mathcal{W}_t)$

① Prediction :

- Calculer  $\mu_t^i = \mathbb{E}(x_t | x_{t-1}^i, \theta_{t-1}^i)$  et  $m_{t-1}^i = \mathbb{E}[\theta_t | \theta_{t-1}^i] = \theta_{t-1}^i$
- Poser  $w_{t-1}^i = \tilde{w}_{t-1}^i p(y_t | \mu_t^i, m_{t-1}^i)$ .

② Rééchantillonnage :

③ Propager :

- Poser  $x_{t-1}^i = \tilde{x}_{t-1}^i$  et  $\theta_{t-1}^i = \theta_{t-1}^i$ .
  - Simuler  $\theta_t^i \sim \mathcal{N}(m_{t-1}^i, \mathcal{W}_t)$
  - Simuler  $x_t^i \sim p(x_t | x_{t-1}^i, \theta_t^i)$ .
  - Calculer les poids :  $w_t = p(y_t | x_t^i, \theta_t^i) / p(y_t | \mu_t^i, m_{t-1}^i)$  et normaliser  $\tilde{w}_t$ .
- 
- **Idée** : Propager de manière à avoir une variance constante.
  - **Problème** :  $\text{Var}(\theta_t) = \text{Var}(\theta_{t-1}) + \mathcal{W}_t$  : Les variances s'accumulent ! Perte d'information !

## Algorithme :

- **Pour**  $t = 1$  **et for**  $i = 1 \cdots M$  :
  - Simuler  $\theta_1^i \sim p(\theta_1)$  et  $x_1^i \sim p(x_1|\theta_1^i)$
  - Calculer  $\tilde{w}_1^i = \frac{1}{M}$ .
- **Pour**  $t = 1, \dots, n$ 
  - 1 **Pour**  $i = 1, \dots, M$  :
    - Calculer  $\mu_t^i = \mathbb{E}(x_t|x_{t-1}^i, \theta_{t-1}^i)$  et  $m_{t-1}^i = a\theta_{t-1}^i + (1-a)\bar{\theta}_{t-1}$ .
    - Poser  $w_{t-1}^i = \tilde{w}_{t-1}^i p(y_t|\mu_t^i, m_{t-1}^i)$ .
  - 2 **Rééchantillonner** :
    - Rééchantillonner  $\{x_{t-1}^i, w_{t-1}^i\}$
    - On obtient  $\{\tilde{x}_{t-1}^i, 1/M\}$  et un échantillon d'indices  $\{j^i\}$ .
  - 3 **Propager** :
    - Poser  $x_{t-1}^i = \tilde{x}_{t-1}^i$  et  $\theta_{t-1}^i = \theta_{t-1}^i$ .
    - Simuler  $\theta_t^i \sim \mathcal{N}(m_{t-1}^i, (1-a^2)V_{t-1})$ .
    - Simuler  $x_t^i \sim p(x_t|x_{t-1}^i, \theta_t^i)$ .
    - Calculer les poids :  $w_t = \frac{p(y_t|x_t^i, \theta_t^i)}{p(y_t|\mu_t^i, m_{t-1}^i)}$  et normaliser  $\tilde{w}_t$ .

Modèle :

$$\begin{cases} dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t = \kappa(\beta - V_t) dt + \sigma \sqrt{V_t} dW_t^2 \\ \langle dW_t^1, dW_t^2 \rangle = \rho dt \end{cases}$$

- $S_t$  : actif et  $r$  le taux sans risque.
- $W_t^1$  et  $W_t^2$  sont deux mouvements browniens.
- $\kappa$  : paramètre de vitesse de retour à la moyenne.
- $\beta$  : moyenne à long terme du processus.
- $\sigma$  : paramètre de "Vol de Vol".
- $\rho$  : paramètre de corrélation.
- $\theta = (\kappa, \beta, \sigma, \rho)$

Modèle populaire utilisé en finance :

- 1 représente les faits observés empiriquement : volatilité clustering, effet de levier...
- 2 formule semi fermée des prix d'options notés  $C_t(v_t, S_t, \theta)$ .

Modèle :

$$\begin{cases} y_t = C_t(v_t, S_t, \theta) + \varepsilon_t, \\ v_t = \Psi(v_{t-1}, \theta, \Delta) + \Phi^{1/2}(v_{t-1}, \theta, \Delta)\eta_t. \end{cases}$$

avec :



$$\begin{cases} \Psi(v_t, \theta, \Delta) = \beta(1 - \exp^{-\kappa\Delta}) + \exp^{-\kappa\Delta}v_t, \\ \Phi(v_t, \theta, \Delta) = \beta\frac{\sigma^2}{2\kappa}(1 - e^{-\kappa\Delta})^2 + \frac{\sigma^2}{\kappa}e^{-\kappa\Delta}(1 - e^{-\kappa\Delta})v_t. \end{cases}$$

- $\Delta$  : pas de temps.
- $\varepsilon_t \sim \mathcal{N}(0, R)$  où  $R$  est connu.

- **Notation** :  $C_t(S_t, v_t, \theta)$  représente le vecteur contenant les prix d'options à la date  $t$  :

$$C_t(S_t, v_t, \theta) = \begin{pmatrix} C_t^{K_1, T_1}(S_t, v_t, \theta) \\ C_t^{K_1, T_2}(S_t, v_t, \theta) \\ \vdots \\ C_t^{K_3, T_3}(S_t, v_t, \theta) \end{pmatrix}$$

- **But** : Estimer séquentiellement et simultanément les paramètres et la variance.

## Paramètres des Filtres :

- $p(\theta_0) = \mathcal{U}(\kappa_a, \kappa_b) \otimes \mathcal{U}(\beta_a, \beta_b) \otimes \mathcal{U}(\sigma_a, \sigma_b) \otimes \mathcal{U}(\rho_a, \rho_b)$
- $p(v_0|\theta_0) = \Gamma(\alpha_1, \alpha_2)$  avec  $\alpha_1 = \frac{2\kappa_0\beta_0}{\sigma_0^2}$  et  $\alpha_2 = \frac{\sigma_0^2}{2\kappa_0}$ .
- $p(v_t|v_{t-1}, \theta_{t-1}) = 2c\chi^2(2d + 2, 2w)$  avec :

$$c = \frac{2\kappa}{\sigma^2(1 - \exp^{-\kappa\Delta})}, \quad w = cv_{t-1}\exp^{-\kappa\Delta} \text{ and } d = \frac{2\kappa\beta}{\sigma^2} - 1$$

- $p(y_t|v_t, \theta_t) = \mathcal{N}(C_t(S_t, v_t, \theta_t), R)$ .

# Résultat numérique : $\theta = (4, 0.03, 0.4, -0.5)$

TABLE: Estimation des paramètres par EKF, APF, APFS et KSAPF. Le nombre de particules  $M$  pour les filtres particulaires est 1000 et le nombre d'observation  $T = 100$ .

	$\kappa_{mean}$	$\beta_{mean}$	$\sigma_{mean}$	$\rho_{mean}$
APF	4.15	0.031	0.46	-0.43
KSAPF	4.05	0.0296	0.41	-0.59
APFS	4.55	0.028	0.43	-0.52
EKF	4.21	0.0296	0.345	-0.61

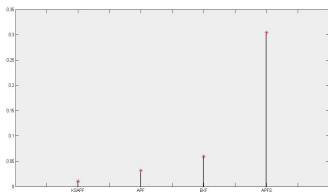
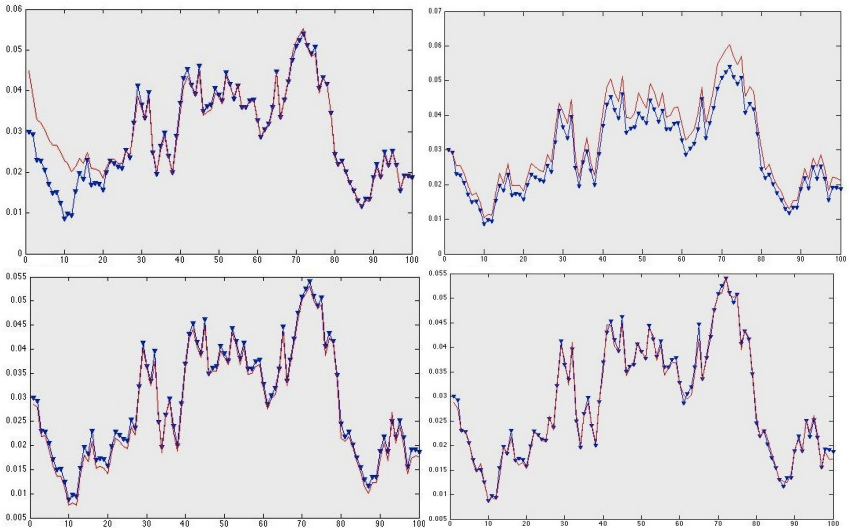
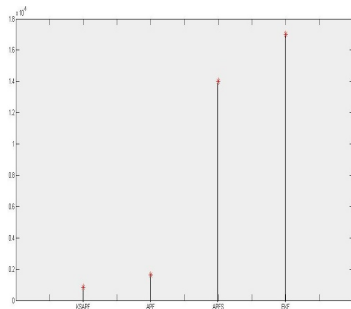


FIGURE: MSE pour l'estimation de  $\theta$ . De gauche à droite : KSAPF, APF, EKF, APFS



**FIGURE:** Estimation de la variance : de gauche à droite : EKF, APFS, APF, KSAPF. Bleu : vraie variance, Rouge : variance estimée

# Résultat numérique :



**FIGURE:** MSE pour l'estimation de  $v_t$ . De gauche à droite : KSAPF, APF, APFS, EKF

- Mauvaise performance du filtre de Kalman Etendu (prix du call fortement non-linéaire) et du filtre APFS (dégénérescence des particules).
- KSAPF performe les autres filtres dans tous les cas
- Paramètre “tuné” pour l’APF : la variance des bruits pour la dynamique artificielle des paramètres.