

Contents

1	Outil mathématique	3
1.1	Pocessus stochastique	3
1.1.1	Distributions de probabilité	4
1.1.2	stationnarité	4
1.1.3	Ergodicité	6
1.2	Filtre de kalman	8
1.3	Analyse Bayésienne et Chaînes de Markov	15
1.3.1	Introduction	15
1.3.2	Principe de l'analyse Bayésienne	16
1.3.3	Chaînes de Markov Discrètes	19
1.4	Aperçu sur les models de volatilité	22
1.4.1	Introduction	22
1.4.2	Les Models ARCH	22
2	Model de volatilité stochastique	26
2.1	Introduction	26
2.2	Présentation du model	27
2.3	Propriétés probabilistes du modèle	29
2.3.1	Stationnarité stricte	29
2.3.2	Stationnarité au second ordre	30
3	Estimation du model de volatilité stochastique	32
3.1	Méthode du Quasi-Maximum de vraisemblance	32
3.1.1	Introduction	32
3.1.2	Présentation de la méthode	32

3.2	Méthode de Monte-Carlo par Chaîne de Markov	41
3.2.1	Introduction	41
3.2.2	Présentation de la méthode	42
3.2.3	Estimation des paramètres	45

1 Outil mathématique

1.1 Processus stochastique

Une série chronologique est la réalisation d'un processus aléatoire $\{Y_t\}$ indexé par le temps. L'étude des séries chronologiques a comme objectifs ; de comprendre le passé de la série, prédire les valeurs futuresetc. La majorité des séries chronologiques sont exprimées par des équations aux différences stochastiques avec, un processus d'entrée ou le plus souvent c'est un processus d'innovation, la forme de l'équation avec ces paramètres, et une sortie qui représente une solution de cette équation qui est elle-même un processus stochastique. L'étude des séries chronologiques revient à étudier la structure des propriétés de sortie à savoir ; la stationnarité et l'ergodicité, en fonction des paramètres et de la forme de l'équation.

Définition 1 *Un processus aléatoire de domaine d'évolution T , défini sur un espace de probabilité (Ω, A, P) et à valeurs dans un espace d'état E et muni d'une tribu ξ ; est une famille de variables aléatoires $(Y_t; t \in T)$ chacune définie sur le même espace (Ω, A, P) et à valeur dans $(E; \xi)$*

Les processus aléatoires peuvent être classés selon la dénombrabilité ou non des espaces d'états E et du domaine d'évolution T . On peut classer les processus aléatoires dans le tableau suivant selon la dénombrabilité ou non de leurs espaces d'état et leurs domaines d'évolution

$T \backslash E$	Dénombrable	Non dénombrable
Dénombrable	Processus à temps discret et a espace d'état discret	Processus à temps discret et a espace d'état continu
Non dénombrable	Processus à temps continu et a espace d'état discret	Processus à temps continu et a espace d'état continu

Dans la suite on s'intéresse au cas des processus à temps discret ($T = \mathbb{Z}$ où $T = \mathbb{N}$) et à espace d'état fini ou infini

1.1.1 Distributions de probabilité

La distribution de probabilité d'un processus aléatoire $\{Y_t, t \in \mathbb{Z}\}$, définie par les loi de toute sous-famille fini $(Y_{t_1}, \dots, Y_{t_i}, \dots, Y_{t_n})$, $i = 1 \dots n$ où $n \in \mathbb{N}^*$, est souvent déterminée à travers ça fonction de répartition

Définition 2 *La fonction de répartition d'un processus aléatoire $\{Y_t, t \in \mathbb{Z}\}$ est définie par*

$$\begin{cases} F_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}}(y_1, y_2, \dots, y_n) = P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2, \dots, Y_{t_n} \leq y_n) \\ \text{pour tout } n \in \mathbb{N}^*, \dots, t_1, t_2, \dots, t_n \in \mathbb{Z} \quad , \quad (y_1, y_2, \dots, y_n) \in \mathbb{R}^n \end{cases}$$

1.1.2 stationnarité

La propriété de stationnarité (stochastique), qui caractérise plutôt une certaine régularité stochastique dans l'évolution, joue un rôle crucial dans la théorie des processus aléatoires. Dans plusieurs problèmes du monde réel, on rencontre des phénomènes aléatoires qui évoluent dans un régime "d'équilibre stochastique" dans le sens où les caractéristiques fréquentistes du phénomènes

ne changent pas dans le domaine d'évolution. De tels phénomènes peuvent être représentés par les dits "processus stationnaires" dont on distingue deux types.

Définition 3 (*stationnarité stricte*)

un processus aléatoire $(Y_t; t \in \mathbb{Z})$ définie sur espace (Ω, A, P) et à valeur dans $(\mathbb{R}; B(\mathbb{R}))$, est dit strictement stationnaire si $\forall t_i, h \in \mathbb{Z} \quad \forall n \in \mathbb{N}^* \quad \forall y_i \in \mathbb{R}$

$$P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2, \dots, Y_{t_n} \leq y_n) = P(Y_{t_1+h} \leq y_1, Y_{t_2+h} \leq y_2, \dots, Y_{t_n+h} \leq y_n)$$

tel que $i = 1, \dots, n$ et $t_{i+h} \in \mathbb{Z}$.

Autrement dit, la distribution de probabilité est invariante par translation dans le domaine d'évolution

Tout les moments d'un processus strictement stationnaire si ils existent, sont invariants dans le temps, c'est pourquoi cette définition reste très exigeante. Plusieurs propriétés probabilistes essentielles des processus aléatoires peuvent être obtenues juste à partir des deux premiers moments (lorsqu'ils existent). La stationnarité faible est un concept de stationnarité moins fort mais qui est très utilisé en pratique.

Définition 4 (*stationnarité faible*)

Un processus aléatoire $(Y_t; t \in \mathbb{Z})$ définie sur espace (Ω, A, P) et à valeur dans $(\mathbb{R}; B(\mathbb{R}))$, est dit faiblement stationnaire (stationnaire au 2eme ordre) si

i) La fonction variance $\sigma^2(t) = \text{Var}(Y_t) < \infty$ existe pour tout t dans le domaine d'évolution \mathbb{Z} (donc nécessairement $\mu(t) = E(Y_t)$ existe : $E(Y_t^2) < \infty$)

ii) La fonction moyenne $\mu(t)$ est constante sur le domaine d'évolution et la fonction d'autocovariance $\gamma(Y_t, Y_{t+h}) = \text{cov}(Y_t, Y_{t+h})$ dépend seulement de h ce qui entraîne une fonction de variance $\sigma^2(t)$ constante dans le domaine d'évolution \mathbb{Z}

Le processus faiblement stationnaire le plus simple est le processus bruit blanc

Définition 5 un processus $(\varepsilon_t, t \in T)$ est dit bruit blanc si est seulement si

- i) $E[\varepsilon_t] = 0$
- ii) $E[\varepsilon_t^2] = \sigma^2$
- iii) $E[\varepsilon_t \varepsilon_h] = 0 \quad \forall t \neq h$

On distingue deux types de bruit blanc :

Un **Bruit blanc fort** est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne zéro et de variance finie.

Un **Bruit blanc faible** est une suite de variables aléatoires non-corrélées de moyenne zéro et de variance finie.

1.1.3 Ergodicité

L'ergodicité exprime le fait que le processus peut prendre n'importe quelle valeur dans l'espace des états indépendamment de sa valeur initiale, et qu'il ne peut s'isoler dans un sous ensemble strictement inclus dans l'espace des états. C'est une propriété importante des processus stationnaires pour l'établissement du théorème ergodique.

Le théorème ergodique représente une généralisation du théorème des grands nombres au cas des processus non *i.i.d.* mais stationnaires. Ce théorème stipule que sous certaines conditions, la moyenne empirique d'une fonction du

processus converge vers une certaine variable aléatoire. Lorsque le processus est ergodique cette variable aléatoire si elle existe, représente la moyenne commune de tout les membres.

Donnons d'abord le concept d'invariance d'un espace par rapport à une transformation. avant de définir la notion d'ergodicité

Définition 6 soit $\mathbb{R}^{\mathbb{Z}}$ l'espace de toutes les séquences $(Y_t, t \in \mathbb{Z})$ de nombres réels

i) l'application $\Upsilon : \mathbb{R}^{\mathbb{Z}} \longrightarrow \mathbb{R}^{\mathbb{Z}}$ est dite opérateur de translation si

$$\Upsilon(..., Y_0, Y_1, Y_2, ...) = (... , Y_1, Y_2, Y_3, ...)$$

ii) Un sous ensemble $A \subset \mathbb{R}^{\mathbb{Z}}$ est dit Υ -invariant si $\Upsilon(A) = A$

Définition 7 (Ergodicité)

Un processus stochastique $(Y_t, t \in \mathbb{Z})$ stationnaire est dit ergodique si

$$P((Y_1, Y_2, Y_3, ...) \in A) = 0 \text{ ou } 1$$

pour chaque ensemble A Υ -invariant.

Lorsque le processus $(Y_t, t \in \mathbb{Z})$ est strictement stationnaire, la suite définie par $\frac{1}{n} \sum_{t=1}^n Y_t$ converge vers une variable aléatoire, mais si de plus le processus $(Y_t, t \in \mathbb{Z})$ est ergodique, une telle variable n'est rien d'autre que la moyenne (une constante). d'où le théorème ergodique suivant

Théorème 8 soit $(Y_t, t \in \mathbb{Z})$ une processus ergodique strictement stationnaire avec une moyenne fini m alors :

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n Y_t = m\right) = 1$$

On peut consulter le théorème *Karlin et Taylor(1975)* pour voir les différentes caractérisations de l'ergodicité

1.2 Filtre de kalman

Le filtre de kalman a été développé originairement par *Kalman (1960)* dans le contexte des systèmes linéaires. Cependant, étant donné la facilité de mise en oeuvre de l'algorithme sur les calculateurs numériques, il est devenu largement utilisé dans de nombreux domaines d'applications en particulier les applications statistiques incluant les prévisions des chroniques. Le filtre de Kalman est une méthode de prévision bayésienne, c'est un estimateur-prédicteur très puissant, qui peut être appliqué sur des modèles univariés ou multivariés. Il est également décrit comme une procédure récursive qui calcule à un instant t l'estimateur optimal de la variable d'état d'une façon séquentielle utilisant la technique du maximum de vraisemblance.

Considérons le modèle espace d'états linéaire suivant

$$\begin{cases} x_{t+1} = F_t x_t + \varepsilon_{t+1} \\ y_t = H_t x_t + A_t u_t + e_t \end{cases}$$

Où $\{x_t\}$, $\{y_t\}$, $\{e_t\}$ et $\{\varepsilon_t\}$ sont des processus aléatoires de dimensions $r \times 1$, $m \times 1$, $r \times 1$ et $m \times 1$ respectivement, tel que

$$\begin{cases} E(e_t) = E(\varepsilon_t) = 0, & E(e_t x'_{t-k}) = 0 \\ E(\varepsilon_t \varepsilon'_t) = \delta_{h,0} Q_t, & E(e_t e'_{t+h}) = \delta_{h,0} R_t \end{cases} \quad \forall t, h \in \mathbb{Z}, \forall k \geq 0$$

Où δ désigne la fonction de Kronecker. On supposant que les matrices F_t , R_t , Q_t et H_t sont connues, la prévision optimale de x_t , sur la base de l'information disponible jusqu'à l'instant $t - 1$, est déviée d'un résultat bien connu pour les vecteur gaussien suivant :

Proposition 9 Soit x_1, x_2 deux vecteurs gaussiens avec

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$$

Par conséquent la distribution de x_2 conditionnée par x_1 suit $N(m; \sum)$ avec

$$\begin{aligned} m &= \mu_2 + \Omega_{21}\Omega_{11}^{-1}(x_1 - \mu_1) \\ \sum &= \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{aligned}$$

Le filtre de Kalman est un algorithme permettant de :

- i) Prédire la valeur du vecteur d'état à la date t sachant ses valeurs passées
- ii) Filtrer, c.à.d. estimer la valeur de x_t à partir des observations de y jusqu'à la t
- iii) Lisser, c.à.d. estimer la valeur de x_t à partir des observations de y jusqu'à la date T , avec $T > t$

Comme première étape (étape de prédiction) notre objectif est d'estimer le vecteur d'état x_t à partir de l'information disponible à l'instant $t - 1$ de façon optimal et recursive sous les hypothèses suivantes

H_1 : le bruit $\{\varepsilon_t\}$ est un bruit blanc gaussien de matrice de variance covariance Q_T .

H_2 : le bruit d'observation $\{e_t\}$ est un bruit blanc gaussien de matrice de variance-covariance R_t .

H_3 : le vecteur initial x_0 est gaussien de moyenne μ_0 et de matrice de variance-covariance P_0 .

H_4 : les bruits $\{\varepsilon_t\}$, $\{e_t\}$ et l'état initial x_0 sont mutuellement indépendants.

et supposons également, que l'information disponible à l'instant $t - 1$ peut être récapitulée dans le vecteur $I_{t-1} = (y'_{t-1}, y'_{t-2}, \dots, u'_{t-1}, u'_{t-2}, \dots)'$.

Si nous adoptons le critère de l'erreur moyenne quadratique, il s'agit de calculer la loi conditionnelle du vecteur d'état x_t sachant I_{t-1} , et d'après les hypothèses H_1, H_3 il suffit de calculer la moyenne $\hat{x}_{t|t-1}$, et la matrice de variance-covariance $P_{t|t-1}$ de cette loi conditionnelle.

Sous les hypothèses H_1, H_4 le vecteur $(y'_0, x'_0)'$ est gaussien de moyenne et de

matrice variance covariance donnée par

$$\left(\begin{bmatrix} H_0\mu_0 + A_0u_0 \\ \mu_0 \end{bmatrix}, \begin{bmatrix} H_0P_0H'_0 + R_0 & H_0P_0 \\ P_0H_0 & P_0 \end{bmatrix} \right)$$

d'où la loi de x_0 sachant y_0 est gaussienne de moyenne

$$\hat{x}_{0/0} = \mu_0 + P_0H'_0(H_0P_0H'_0)^{-1}(y_0 - H_0\mu_0 - A_0u_0)$$

et de matrice de variance-covariance

$$P_{0/0} = P_0 - P_0H'_0(H_0P_0H'_0)^{-1}H_0P_0$$

Supposons maintenant que nous sommes à l'instant t et nous connaissons les valeurs de $\hat{x}_{t-1/t-1}$ et $P_{t-1/t-1}$ qui caractérise la distribution conditionnelle de x_{t-1} sachant I_{t-1} . Le vecteur $(x'_t, y'_1, \dots, y'_{t-1})'$ est gaussien, donc la loi de x_t sachant I_{t-1} est gaussienne, de moyenne $\hat{x}_{t/t-1}$ et de matrice de variance-covariance $P_{t/t-1}$, nous avons alors :

$$\begin{aligned} \hat{x}_{t/t-1} &= E(x_t/u_t, I_{t-1}) = E(F_{t-1}x_{t-1} + \varepsilon_t/u_t, I_{t-1}) \\ \hat{x}_{t/t-1} &= F_{t-1}\hat{x}_{t-1/t-1} \end{aligned}$$

et par conséquent

$$P_{t/t-1} = E[(x_t - \hat{x}_{t/t-1})(x_t - \hat{x}_{t/t-1})'/u_t, I_{t-1}]P_{t/t-1} = F_{t-1}P_{t-1/t-1}F'_{t-1} + Q_t$$

d'où les equations de prédiction sont données par

$$\begin{aligned} \hat{x}_{t/t-1} &= F_{t-1}\hat{x}_{t-1/t-1} \\ P_{t/t-1} &= F_{t-1}P_{t-1/t-1}F'_{t-1} + Q_t \end{aligned} \tag{1}$$

Dans l'étape du filtrage, l'objectif est de calculer la loi conditionnelle du vecteur d'état x_t sachant I_t , alors que celle de x_t sachant I_{t-1} est connue. Nous tenons compte de la nouvelle information par rapport à I_{t-1} apportée par $(y'_t, u'_t)'$, i.e

$$\hat{e}_t = y_t - \hat{y}_t = y_t - E(y_t/u_t, I_{t-1}) = H_t(x_t - \hat{x}_{t/t-1}) + e_t$$

où le processus $\{\hat{e}_t\}$ est un processus gaussien (processus d'innovation) de moyenne nulle et de matrice de variance-covariance :

$$E[(y_t - \hat{y}_t)(y_t - \hat{y}_t)'/u_t, I_{t-1}] = H_t P_{t/t-1} H_t' + R_t$$

Dans cette dernière égalité nous utilisons le fait que $(x_t - \hat{x}_{t/t-1})$ est indépendant de e_t et par conséquent

$$E[(x_t - \hat{x}_{t/t-1})e_t'/u_t, I_{t-1}] = E[(x_t - \hat{x}_{t/t-1})'e_t/u_t, I_{t-1}] = 0$$

Il est clair que le vecteur $(x'_t, y'_1, \dots, y'_t)'$ est gaussien, ainsi la loi de x_t sachant I_t est gaussienne, de moyenne $\hat{x}_{t/t}$ et de matrice variance-covariance $P_{t/t}$. Nous avons

$$\hat{x}_{t/t} = E(x_t/I_t) = \hat{x}_{t/t-1} + E[(x_t - \hat{x}_{t/t-1})/I_t]$$

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + E[(x_t - \hat{x}_{t/t-1})/I_{t-1}, u_t, \hat{e}_t]$$

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + E[(x_t - \hat{x}_{t/t-1})/\hat{e}_t]$$

d'où

$$x_t - \hat{x}_{t/t} = (x_t - \hat{x}_{t/t-1}) - (x_t - \hat{x}_{t/t-1})$$

$$x_t - \hat{x}_{t/t} = (x_t - \hat{x}_{t/t-1}) + E[(x_t - \hat{x}_{t/t-1})/\hat{e}_t]$$

par conséquent

$$P_{t/t} = E[(x_t - \hat{x}_{t/t})(x_t - \hat{x}_{t/t})']$$

$$P_{t/t} = E[((x_t - \hat{x}_{t/t-1}) - E[(x_t - \hat{x}_{t-1})/\hat{e}_t]) \times ((x_t - \hat{x}_{t/t-1}) - E[(x_t - \hat{x}_{t-1})/\hat{e}_t])]'$$

Il suffit donc de calculer la moyenne conditionnelle et la matrice de variance-covariance conditionnelle du vecteur $(x_t - \hat{x}_{t/t-1})$ sachant \hat{e}_t . A cette fin, nous savons que le vecteur $((x_t - \hat{x}_{t-1})', \hat{e}_t')'$ est un vecteur gaussien, de moyenne nulle et de matrice de variance-covariance

$$\begin{pmatrix} P_{t/t-1} & P_{t/t-1}H_t' \\ H_tP_{t/t-1} & H_tP_{t/t-1}H_t' + R_t \end{pmatrix}$$

de plus en supposant que la matrice R_t est inversible, alors $H_tP_{t/t-1}H_t' + R_t$ est inversible aussi et nous avons immédiatement les equations de mises à jours

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + P_{t/t-1}H_t'[H_tP_{t/t-1}H_t' + R_t]^{-1}\hat{e}_t \quad (2)$$

$$P_{t/t} = P_{t/t-1} - P_{t/t-1}H_t'[H_tP_{t/t-1}H_t' + R_t]^{-1}H_tP_{t/t-1}$$

Le filltre de Kalman est formé par les équations de 1 et 2., ou il a une structure dite "prediction-correction", ou "prédiction-mise à jour". On commence par former une anticipation, puis on met à jour cette prévision, et ceci de façon récursive. En récapitulation; le filtre de kalman pour les models espaces d'états linéaire, est un algorithme de calcul séquentiel de $\{\hat{x}_{t+1/t}\}_{t=1}^T$ et de $\{P_{t+1/t}\}_{t=1}^T$

Algorithme 10

$$\Omega_t = H_tP_{t/t-1}H_t' + R_t$$

$$K_t = F_tP_{t/t-1}H_t\Omega_t^{-1}$$

$$\hat{x}_{t+1/t} = (F_t - K_tH_t)\hat{x}_{t/t-1} + K_t y_t - K_t A_t u_t$$

$$P_{t+1/t} = F_tP_{t/t-1}F_t' - K_t\Omega_t K_t' + Q_t$$

où K_t est appelée *gain de kalman* .

Il peut être intéressant de recalculer recursivement la variable d'état (en arrière) une fois que toute l'information jusqu'à l'instant t est révélée. On appelle cette opération **le lissage**. On définit alors

$$\begin{aligned}\hat{x}_{t/T} &= E[x_t/I_T] & \forall T > t \\ P_{t/T} &= E[(x_t - \hat{x}_{t/T})(x_t - \hat{x}_{t/T})']\end{aligned}$$

Supposons que la valeur réelle de x_t soit connue, à partir de la formule de mise à jour de la projection linéaire de t sur l'ensemble des observations déjà utilisée, permettant de calculer le nouvel estimateur de x_t formulé comme suit

$$E[x_{t-1}/x_t] = x_{t-1/t-1} + E[(x_{t-1} - x_{t-1/t-1})(x_t - x_{t/t-1})'] E[(x_t - \hat{x}_{t/t-1})(x_t - \hat{x}_{t/t-1})']^{-1} (x_t - \hat{x}_{t/t-1}) \quad (1)$$

la première espérance s'écrit comme suite :

$$E[(x_{t-1} - x_{t-1/t-1})(x_t - x_{t/t-1})'] = E[(x_{t-1} - x_{t-1/t-1})(F_t x_{t-1} + \varepsilon_t - F_t x_{t-1/t-1})]$$

le fait que ε_t soit non corrélés avec x_{t-1} et $x_{t-1/t-1}$ implique que

$$E[(x_{t-1} - x_{t-1/t-1})(x_t - x_{t/t-1})'] = E[(x_{t-1} - x_{t-1/t-1})(x_{t-1} - x_{t-1/t-1})'] F_t'$$

en remplaçant dans 5 on obtient

$$E[x_{t-1}/x_t] = x_{t-1/t-1} + J_{t-1}(x_t - \hat{x}_{t/t-1})$$

où J_{t-1} est la matrice de gain du problème de lissage tel que

$$J_{t-1} = P_{t-1/t-1} F_t' P_{t/t-1}^{-1}$$

la suite des estimations lissée de $\{x_t\}_{t=1}^T$ est calculée comme suit

$$x_{t/T} = x_{t/t} + J_t(x_{t+1/T} - x_{t+1/t}) \quad (2)$$

Le déroulement de l'algorithme du filtre de Kalman nous a permis de déterminer les valeurs des suite $\{x_{t/t}\}_{t=1}^T$, $\{x_{t/t-1}\}_{t=2}^T$, $\{P_{t/t}\}_{t=1}^T$, $\{P_{t/t-1}\}_{t=2}^T$ lesquelles sont stockées respectivement dans chaque terme de ces suites. L'estimation lissée de la dernière valeur de l'échantillon $x_{T/T}$ est représentée par la dernière entrée dans la suite $\{x_{t/t}\}_{t=1}^T$.

soit $t = T - 1$ et $\{J_t\}_{t=1}^T$ comme générateur alors l'équation 2

$$x_{T-1/T} = x_{T-1/T-1} + j_{T-1}(x_{T/T} - x_{T/T-1})$$

et de la même manière en posant $t = (T - 2), (T - 3), \dots$, on détermine toutes les valeurs lissées à partir de l'échantillon $\{x_{t/T}\}_{t=1}^T$.

Afin de déterminer la matrice de covariance de l'erreur associée à l'estimation de lissage nous calculons d'abord $(x_t - x_{t/T})$ de l'équation 2

$$\begin{aligned} x_t - x_{t/T} &= x_t - x_{t/t} - J_t(x_{t+1/T} - x_{t+1/t}) \\ x_t - x_{t/T} + J_t x_{t+1/T} &= x_t - x_{t/t} + J_t x_{t+1/t} \end{aligned}$$

En prenant l'espérance des deux membres multiplier par leurs transposé

$$E[(x_t - x_{t/T} + J_t x_{t+1/T})(x_t - x_{t/T} + J_t x_{t+1/T})'] = E[(x_t - x_{t/t} + J_t x_{t+1/t})(x_t - x_{t/t} + J_t x_{t+1/t})']$$

et en simplifiant par la non corrélation entre $(x_t - x_{t/T}), (x_{t+1/T}), (x_t - x_{t/t})$

et $x_{t+1/t}$ on a

$$E[(x_t - x_{t/T})(x_t - x_{t/T})'] + J_t E(x_{t+1/T} x_{t+1/T}') J_t' = E[(x_t - x_{t/t})(x_t - x_{t/t})'] + J_t E(x_{t+1/t} x_{t+1/t}') J_t'$$

d'où

$$P_{t/T} = P_{t/t} + J_t \{-E(x_{t+1/T} x_{t+1/T}') + E(x_{t+1/t} x_{t+1/t}')\} J_t' \quad (3)$$

tel que $\{-E(x_{t+1/T}x'_{t+1/T}) + E(x_{t+1/t}x'_{t+1/t})\}$ s'écrit comme suit :

$$\begin{aligned} & \{-E(x_{t+1}x'_{t+1}) - E(x_{t+1/T}x'_{t+1/T})\} - \{-E(x_{t+1}x'_{t+1}) - E(x_{t+1/t}x'_{t+1/t})\} \\ & E[(x_{t+1} - x_{t+1/T})(x_{t+1} - x_{t+1/T})'] - E[(x_{t+1} - x_{t+1/t})(x_{t+1} - x_{t+1/t})'] \end{aligned}$$

d'où

$$\{-E(x_{t+1/T}x'_{t+1/T}) + E(x_{t+1/t}x'_{t+1/t})\} = P_{t+1/T} - P_{t+1/t}$$

En remplaçant dans l'équation 3

$$P_{t/T} = P_{t/t} + J_t(P_{t+1/T} - P_{t+1/t})J'_t$$

La détermination des matrices de variance-covariance associées à chaque estimateur lissée s'obtient de la même manière que précédemment, en commençant par $t = (T - 1), (T - 2) \dots$

1.3 Analyse Bayésienne et Chaînes de Markov

1.3.1 Introduction

En statistique bayésienne particulièrement, mais aussi en statistique fréquentiste, on a besoin d'intégrer par rapport à des lois de probabilités multivariées pour faire des inférences sur les paramètres de notre modèle d'étude, où pour faire des prédictions. Quand les calculs deviennent impraticables on peut avoir recours à un ordinateur pour produire des simulations. Les capacités des ordinateurs actuels permettent de programmer des algorithmes *MCMC* (Markov Chain Monte Carlo) dont l'existence remonte à 1953, qui simulent des trajectoires d'une chaîne de Markov dont la distribution empirique approche la loi désirée.

L'analyse bayésienne repose sur des concepts clés à savoir, la densité a priori, la densité a posteriori et la proportionnalité qu'on note par \propto . Nous présentons les fondements de la théorie bayésienne dans une première partie, la deuxième partie sera consacrée au principe des chaînes de Markov dans un espace d'état discret.

1.3.2 Principe de l'analyse Bayésienne

Soit $Y = (Y_1, \dots, Y_n)$ un n-échantillon de variables aléatoires *i.i.d* de densité $f(y/\theta)$ qui dépend d'un paramètre inconnue θ , la vraisemblance des observations y_1, \dots, y_n est notée $L(\underline{y}/\theta)$ où \underline{y} désigne le n-uples (y_1, \dots, y_n)

Dans l'approche bayésienne, l'idée de base consiste à traiter le paramètre inconnu θ comme une variable aléatoire admettant une densité de probabilité $\pi(\theta)$, et afin de l'estimer on combine l'information résumé dans la loi dite a **priori** $\pi(\theta)$ à celle tirée de l'échantillon. L'inférence bayésienne est alors fondée sur une loi a **postérieur** $\pi(\theta/Y)$ qui est une fonction de vraisemblance $L(Y/\theta)$ du model et de loi a priori $\pi(\theta)$.

Définition 11 *Un modèle statistique bayésien est la double donnée d'un modèle paramétrique $\{f_\theta(y), \theta \in \Theta\}$ et une loi de probabilité, de densité π , dite loi a priori qui est la loi marginale de la variable aléatoire θ .*

Distribution a propri Le choix de la loi a priori touche d'une façon directe les distributions a posteriori qui sont utilisées dans l'inférence bayésienne. Il y'a deux mode de pensés qui sont mis à l'evidence par rapport à la loi a priori.

Le premier mode de pensé est le cas ou on ne dispose pas d'informations ,sur le

paramètres, on parlera alors de «distribution a priori non informative». Le choix d'une distribution a priori non informative conduit souvent à la spécification d'une mesure et non d'une probabilité dont on distingue trois approches ; la recherche d'a priori invariantes, l'approche de Jeffrey et l'approche dite des a priori de référence, mais dans notre travail on s'intéressera au deuxième mode de pensée

Le deuxième repose sur l'information disponible sur le paramètre obtenue par les chercheurs, elle est exprimée par une loi de probabilité dite «distribution a priori informative». Il existe plusieurs procédés pour obtenir des lois informatives, le plus intéressant est celui des familles naturelles conjuguées

Définition 12 *une famille G de lois a priori pour θ est dite conjuguée pour la vraisemblance $L(\underline{y}/\theta)$ si pour tout $\pi_\theta(\cdot) \in G$, la loi a posteriori $\pi(\theta/Y)$ est un élément de G*

L'approche conjuguée reste la solution la plus standard dans le cadre informatif. Une famille naturelle conjuguée est un concept qui simplifie considérablement le calcul des distributions a posteriori. C'est donc la raison du développement de ce type de lois a priori.

Loi a posteriori soit $Y = (Y_1, \dots, Y_n)$ un n-échantillon de variables aléatoires *i.i.d* de densité $f(y/\theta)$. On suppose de θ est une valeur d'une variable aléatoire Θ de densité $\pi(\theta)$ et soit

$$L(\underline{y}/\theta) = \prod_{i=1}^n f(y_i/\theta)$$

la fonction de vraisemblance des observations. La distribution conditionnelle de θ sachant les observations $y = (y_1, \dots, y_n)$ est alors donnée par :

$$\pi(\theta/\underline{y}) = \frac{L(\underline{y}/\theta)\pi(\theta)}{\int_{\Theta} L(\underline{y}/\theta)\pi(\theta)d\theta} = \frac{\prod_{i=1}^n f(y_i/\theta)\pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(y_i/\theta)\pi(\theta)d\theta}$$

Où encore :

$$\pi(\theta/\underline{y}) \propto \prod_{i=1}^n f(y_i/\theta)\pi(\theta)$$

La distribution a postérieure $\pi(\theta/\underline{y})$ regroupe alors l'information a priori donnée par π et celle tirée de l'échantillon qui est donnée par $L(\underline{y}/\theta)$.

L'inférence bayésienne est fondée sur la détermination rigoureuse de **la loi a priori et a postérieure** et du coût associé à la décision $C(\delta, \underline{y})$. Le choix d'une fonction de coût dépend de l'application concernée. Lorsque le contexte ne conduit pas à la détermination de la fonction d'utilité qui permet d'en déduire la fonction de coût, on peut avoir recours à des fonctions de coûts classiques, qui sont à la fois simples et bien étudiées comme la fonction du coût quadratique :

$$C(\delta, \underline{y}) = \|\theta - \delta^2\|$$

On est amené alors à évaluer un estimateur de θ , qu'on note $\hat{\theta}$ sous une certaine fonction de coût $C(\delta, \underline{y})$

l'estimateur de bayes $\hat{\theta}$ associé à la loi a postérieure $\pi(\theta/\underline{y})$ et au coût quadratique $C(\delta, \underline{y}) = \|\theta - \delta^2\|$ est la **moyenne a posteriori**:

$$\hat{\theta} = E[\theta/\underline{y}] = \int \theta \pi(\theta/\underline{y}) d\theta$$

qui selon *CarlinetLouis(2000, chapitre2)*, est le meilleur estimateur minimisant le coût a postérieure.

1.3.3 Chaînes de Markov Discrètes

Définition 13 un processus stochastique $(S_t, t \in \mathbb{N})$ défini sur un espace de probabilité (Ω, P) à valeurs dans un espace d'état $E = \{1, 2, \dots, K\}$ dénombrable (K peut être infini) est dit chaîne de Markov si pour tout $t \in \mathbb{N}$ et tout $i, j \in E$ ($0 \leq j \leq t-1$)

$$P(S_{t+1} = j | S_t = i, S_{t-1} = i_{t-1}, \dots, S_0 = i_0) = P(S_{t+1} = j | S_t = i) = P_{ij}(t)$$

Cette chaîne est homogène si de plus son évolution ne dépend pas de l'instant t , mais seulement des états concernés ; $P_{ij}(t) = P_{ij} \quad \forall t$.

Définition 14 (Matrice de transition)

Soit $(P_t, t \geq 1)$ une suite de matrices stochastique (ses coefficients sont positifs et la somme sur une ligne des coefficients est égal à 1). On dit que les matrices $(P_t, t \geq 1)$ sont des matrices de transition de la chaîne de Markov S si pour tout $t \geq 1$ et $i \in E$ tels que $P(S_{t-1} = i) > 0$ on a pour tout $j \in E$

$$P_{ij}(t) = P(S_t = j | S_{t-1} = i)$$

La distribution marginale On peut caractériser les distributions marginales d'une chaîne de Markov à partir de la distribution initiale $\pi_j(0) = P(S_0 = j)$ et la matrice de probabilités de transitions P

et on note

$$\pi^n = (\pi_1, \pi_2, \dots, \pi_K)$$

avec

$$\pi_j^n = P(S_n = j), n \in \mathbb{N}, \quad 1 \leq j \leq K$$

Ces probabilités peuvent s'écrire en fonction des probabilités de transition comme suit

$$\begin{aligned}\pi_j &= P(S_n = j) \\ &= \sum_{k \in K} \pi_k P_{kj}^n\end{aligned}$$

qui est équivalente à la forme matricielle suivante

$$\pi = \pi P^n$$

Distribution Stationnaire

Définition 15 Si $\pi_i \geq 0$ pour tout $i \in E$ et $\sum_{i \in E} \pi_i = 1$, alors $(\pi_i, i \in E)$ est appelé la distribution stationnaire (ou invariante)

Le problème fondamental des Chaînes de Markov homogènes est celui de la stabilité qu'on peut résumer à travers les trois questions suivantes :

- i) Existe-il un système de distribution stationnaire (existence)?
- ii) est-il unique si il existe (Unicité)?
- iii) Les distribution marginal convergent-elles vers les distributions stationnaire?

Ces problèmes dépendent de certaines propriétés structurelles de la chaîne comme l'indique le théorème suivant

Théorème 16 Soit $(S_n, n \in \mathbb{N})$ une chaîne de Markov définie sur un espace de probabilité (Ω, A, P) à valeur dans un espace d'état E discret. On suppose que $(S_n, n \in \mathbb{N})$ est

- irréductible
- Apériodique

- *Récurrente positive*

Alors

i) *La distribution stationnaire π existe*

ii) *Elle est unique*

iii) $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j \quad \forall i \in E$

de plus si $\pi(0) = \pi$ alors $(S_n, n \in \mathbb{N})$ est un processus strictement stationnaire

Rappelons que; dans une chaîne de Markov a espace d'état fini E on note $i \rightsquigarrow j$ s'il existe une suite de transitions de probabilité non nulle menant de i à j , et $i \leftrightsquigarrow j$ si $i \rightsquigarrow j$ et $j \rightsquigarrow i$. Un état i tel que pour un j on ait $i \rightsquigarrow j$ et pas $j \rightsquigarrow i$ est dit **transient** ; les autres sont **récurrents**. La relation de communication partitionne l'ensemble des états récurrents en classes d'équivalence appelées classes **d'irréductibilité**. Partant d'un état la chaîne arrive avec probabilité 1 dans une classe d'irréductibilité, en visite chaque état infiniment souvent, et plus jamais les autres états. Une classe d'irréductibilité est dite **périodique** si elle admet une partition fixe, elle est dite **apériodique** sinon.

1.4 Aperçu sur les models de volatilité

1.4.1 Introduction

Dans la modélisation des série chronologique on s'intéresse souvent a la prévision des valeurs future d'une série .L'outil utilisé est l'espérance conditionnelle qui est le meilleur prédicteur au sens du critère de l'erreur quadratique moyenne, dans ce cas on a recours a des modèles linéaires à coefficient constant (*ARIMA, SARIMA.....*) et a coefficients dépendant du temps

(*ARMA périodique, ARMA_t évolutifs...*) qui sont qualifiées de **models de moyenne conditionnelle**, or dans de nombreux domaines scientifique, on s'intéresse plutôt à la variation future des valeurs au lieu des valeurs exactes, on parle alors de «volatilité», dans ce cas les models de moyenne conditionnelle sont limités, car ils ne permettent pas la prise en compte des phénomènes de variabilité future (la variance inconditionnelle et conditionnelle est constante). L'outil de prévision de la volatilité, est la variance conditionnelle. Nous allons donner un aperçu sur les **models de variance conditionnelle** les plus connus en littérature.

1.4.2 Les Models ARCH

La classe de modèle généralement utilisée pour mesurer la volatilité est celle des modèles à **hétéroscédasticité conditionnelle**, plutôt connus sous le nom d'**ARCH** (AutoRegressive Conditional Heteroskedasticity) d'après l'article fondateur d'*Engle (1982)*. Ces models caractérisent la dynamique de la variance conditionnelle que l'on retrouve dans la plupart des séries financières, par la suite, en 1986 *Bollerslev* a généraliser cette classe de models par ce qu'on appelle **Hétéroscédastique Conditionnelle Autorégressive Généralisée ou GARCH**, qui est souvent utilisé lorsque il s'agit d'estimer les variances conditionnelles

Présentation du Model ARCH(p) Le principe général proposé par *Engle* consiste à supposer que la variance dépend de l'ensemble informationnel dont on dispose, il propose une spécification $ARCH(q)$ où le carré des perturbations suit un processus **Autorégressif** d'ordre q

Définition 17 Un processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ satisfait une représentation $ARCH(q)$

si :

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2 \end{cases}$$

Tel que : $\alpha_q \neq 0$ $\alpha_0 > 0$ $\alpha_i \geq 0$ $\forall i = 1, \dots, q$, où (η_t) désigne un bruit blanc faible

On retrouve la propriété de variance conditionnelle dans le temps puisque

$$Var[\varepsilon_t | I_{t-1}] = h_t = \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2$$

Où I_{t-1} est la tribu engendrée par le passé du processus jusqu'au temps $t-1$. Une forme $AR(p)$ est donnée pour les carrés du processus d'innovation, il suffit de poser

$$u_t = \varepsilon_t^2 - h_t \quad (4)$$

d'où :

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2 + u_t$$

avec toutefois la remarque que les variable $\varepsilon_t^2 = (\eta_t^2 - 1)h_t$ ne sont pas de variance constante mais sont de moyenne nulle et non corrélées entre elles.

Plusieurs variantes du modèle ARCH ont été proposées. *Bollerslev (1986)* à généraliser l'effet ARCH en le dotant d'un effet de persistance au niveau de la variance conditionnelle, nous parlons alors de **Généralized ARCH** où encore modèle **GARCH**, qui est l'une des variantes du models ARCH la plus utiliser en littérature

Présentation du Model GARCH(p,q) Dans le modèle **GARCH** la variance conditionnelle h_t dépend aussi bien de ses propres valeurs passées

que des valeurs retardées de ε_t^2 , *Bollerslev(1986)* a donc introduit une dynamique Autoregressive Moyenne Mobile pour la variance conditionnelle par analogie avec la démarche usuelle de *Box et Jenkins (1976)*, permettant ainsi à la structure de retards d'être plus flexible

Définition 18 *Un processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ satisfait une représentation $GARCH(p, q)$ si*

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ Var[\varepsilon_t | I_{t-1}] = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \end{cases}$$

Où (η_t) est un bruit blanc, avec les conditions $\alpha_0 > 0$, $\alpha_i \geq 0$, $\forall i = 1, \dots, q$, $\beta_j \geq 0$, $\forall j = 1, \dots, p$ suffisante pour garantir la positivité de h_t

Tout comme le modèle ARCH, on peut exprimer le processus ε_t^2 sous la forme d'un processus ARMA défini dans une innovation en utilisant la transformation 4 dans l'équation d'un $GARCH(p, q)$ il devient

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j u_{t-j} + u_t$$

Avec la convention $\alpha_i = 0$ si $i > q$ et $\beta_j = 0$ si $j > p$

La classe des modèles ARCH ont certes l'avantage d'être simples mais sont contraints par la forte hypothèse d'une fonction de variance déterministe. En d'autres termes, ces modèles suppose que la variance future est connue de façon certaine ce qui n'est pas tout-à-fait cohérent avec la réalité. Dans le prochain chapitre nous allons introduire une autre classe de modèle, appeler modèles de volatilité stochastique, permettent de mieux capter la volatilité

2 Model de volatilité stochastique

2.1 Introduction

Beaucoup de séries temporelles et surtout les séries financières, comme les rendements d'un actif, ou les taux de change ont une variance qui change dans le temps. on peut essayer de modéliser ces changement de variance de plusieurs façons. Une façon est de représenter la série d'interet comme une suite de bruits blancs avec variance unitaire, multipliés par un facteur d'échelle :

$$y_t = \beta \sqrt{\sigma_t} \varepsilon_t$$

Où ε_t est un bruit blanc gaussien indépendant de y_t , Une possibilité est de modéliser directement la volatilité σ_t en fonction des observations passées de la série, comme dans les models *ARCH*. (où *GARCH*). D'après l'article fondateur **d'Engle 1982**, les modèles *ARCH* reposent sur l'adéquation entre les concepts de variance conditionnelle et de la volatilité, cette spécification, qui fait de la volatilité une variable observable, a d'immenses avantages du point de vue statistique (prévision, inférence), mais elle rend l'étude probabiliste complexe, en d'autres termes, ces modèles suppose que la variance future est observable de façon certaine ce qui n'est pas tout a fait vraie dans la réalité. Une approche alternative consiste à considérer que le logarithme de la volatilité est une variable inobservable suivant processus stochastique, généralement un autoregressif d'ordre 1, ces modèles sont appelés modèles à volatilité stochastique canonique

2.2 Présentation du model

La première apparition du modèle à volatilité stochastique, dans la littérature économétrique a été introduit par *Taylor en 1986*, nous présenterons celui utiliser par *Kim, S, Shepard, N. et Chib, S* en 1998 comme suite :

$$\left\{ \begin{array}{l} y_t = \beta e^{h_t/2} \varepsilon_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \\ h_1 \rightsquigarrow N(\mu, \frac{\sigma_\eta^2}{1-\phi^2}) \end{array} \right. \quad (5)$$

où $|\phi| < 1$, (ε_t) et (η_t) sont des bruits blanc gaussiens mutuellement indépendant

Le processus (Y_t) est la serie temporelle qui est observable et indépendante de (ε_t) .

Le processus (h_t) représente le log-volatilité à l'instant t qui est supposer suivre un processus stationnaire où h_1 est tirées d'une distribution stationnaire.

Le parametre ϕ joue le role de coefficient de persistance, en effet lorsque ϕ est proche de 1, un choc positif sur la volatilité a pour effet de maintenir cette dernière a un niveau élevé sur plusieurs périodes et inversement ,un choc negatif la matien a un niveau bas. Si ϕ est proche de 0 l'effet du choc est transitoir et enfin si il est proche de -1 l'effet instantané d'un choc positif est une volatilité anormalement élevée mais, dès la date suivante, cette volatilité prend une petite valeur, puis à nouveau une grande valeur etc, ceci sous réserve qu'aucun nouveau choc n'intervienne entre temps. Un choc négatif produit le même type d'effets alternés, ceux-ci n'étant généralement pas observés pour les séries financières, on peut d'ores et déjà considérer les valeurs

négatives de ϕ comme peu réalistes pour les applications. L'interprétation des autres coefficients est plus immédiate :

Le paramètre μ où on considère que $\beta = \exp(\mu/2)$ est un facteur d'échelle pour la volatilité.

Le paramètre σ_η qui est constant dans ce modèle, mesure l'amplitude des oscillations de h_t autour de sa moyenne et peut s'interpréter comme la volatilité du log-volatilité.

les figures suivantes représente des simulations du modèle 5

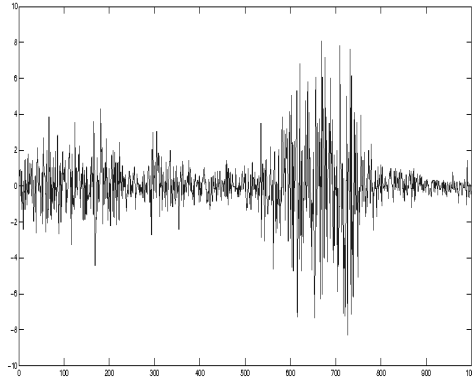


Figure 2.2.1 : $\phi = 0.99, \mu = 0, \sigma^2 = 0.02$

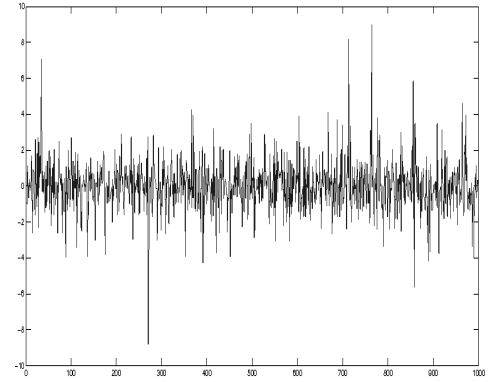


Figure 2.2.2 : $\phi = 0.001, \mu = 0, \sigma^2 = 0.9$

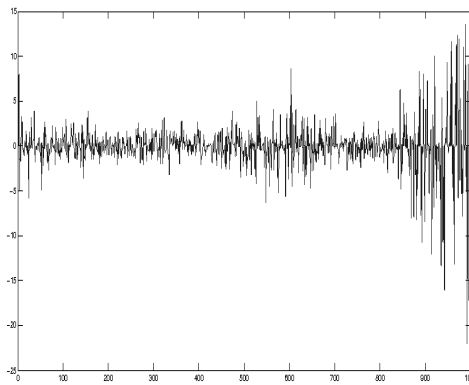


Figure 2.2.3 : $\phi = -0.99, \mu = 0, \sigma^2 = 0.02$

2.3 Propriétés probabilistes du modèle

2.3.1 Stationnarité stricte

Proposition 19 *Le model 5, admet sous la contrainte de $|\phi| < 1$, une unique solution strictement stationnaire qui est non anticipative et donnée par*

$$y_t = \exp \left\{ \mu + \frac{\sigma_\eta}{2} \sum_{i=0}^{\infty} \phi^i \eta_{t-i} \right\} \varepsilon_t \quad , \quad t \in \mathbb{Z} \quad (6)$$

La forme de la solution strictement stationnaire permet de préciser ses caractéristiques du second ordre

Preuve. Pour $\beta = \exp(u/2)$, le model 5 devient

$$\begin{cases} y_t = e^{(h_t + \mu)/2} \varepsilon_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \end{cases}$$

la fonction log volatilité h_t peut s'écrire comme suit

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_{t-1} \quad (7)$$

en remplaçant h_{t-1} dans 7 on a

$$h_t = \mu(1 - \phi) + \phi\mu(1 - \phi) + \phi^2 h_{t-2} + \sigma_\eta \eta_{t-1} + \phi\sigma_\eta \eta_{t-2}$$

En remplaçant encore une autre fois de la même manière on obtient

$$h_t = \mu(1 - \phi) + \phi\mu(1 - \phi) + \phi^2\mu(1 - \phi) + \phi^3 h_{t-3} + \sigma_\eta \eta_{t-1} + \phi\sigma_\eta \eta_{t-2} + \phi^2\sigma_\eta \eta_{t-3}$$

par intuition on peut déduire que

$$\begin{aligned} h_t &= \mu(1 - \phi)[1 + \phi + \phi^2 \dots] + \sigma_\eta [\eta_{t-1} + \phi\eta_{t-2} + \phi^2\eta_{t-3} \dots] \\ h_t &= \mu + \sum_{i=0}^{\infty} \sigma_\eta \phi^i \eta_{t-i-1} \quad \forall t \in \mathbb{Z} \end{aligned} \quad (8)$$

En remplaçant 8 dans 5 on obtien 6 ■

2.3.2 Stationnarité au second ordre

Proposition 20 Notons $\alpha_i = E \{ \exp(\sigma_\eta \phi^i \eta_{t-i-1}) \}, i \geq 0$. Si $\prod_{i=0}^{\infty} \alpha_i < \infty$ et $|\phi| < 1$, alors le processus (y_t) défini par 5 est un bruit blanc de variance :

$$Var(y_t) = \exp(2\mu) \prod_{i=0}^{\infty} \alpha_i$$

Preuve. On a, d'après les hypothèses d'indépendance sur les suites (η_t) et (ε_t) .

$$Var[y_t] = E[y_t^2] - E[y_t]^2 \text{ où } E[y_t] = 0$$

$$\begin{aligned} Var[y_t] &= E \left[\exp \left\{ \mu + \sum_{i=0}^{\infty} \frac{\sigma_\eta}{2} \phi^i \eta_{t-i-1} \right\} \right]^2 E[\varepsilon_t^2] \\ &= \exp(2\mu) \times E \left[\exp \left\{ \sum_{i=0}^{\infty} \sigma_\eta \phi^i \eta_{t-i-1} \right\} \right] \\ Var[y_t] &= \exp(2\mu) \times \prod_{i=0}^{\infty} E \left[\exp \{ \sigma_\eta \phi^i \eta_{t-i-1} \} \right] \quad \forall t \in \mathbb{Z} \end{aligned}$$

sous les hypothèses du model 5 on sait que si η_t est un processus gaussien alors on a le resultat suivant $E[\exp(\alpha \eta_t)] = \exp(\alpha^2/2)$ pour toute constante α d'où

$$\begin{aligned} \prod_{i=0}^{\infty} E[\exp\{\sigma_\eta \phi^i \eta_{t-i-1}\}] &= \exp\left\{ \frac{1}{2} \sum_{i=0}^{\infty} \sigma_\eta^2 \phi^{2i} \right\} \\ &= \exp\left\{ \frac{\sigma_\eta^2}{2(1-\phi^2)} \right\} \end{aligned}$$

d'où

$$Var[y_t] = \exp\left\{ 2\mu + \frac{\sigma_\eta^2}{2(1-\phi^2)} \right\}$$

de plus en utilisant à nouveau l'indépendance entre (ε_t) et (η_t) alors pour tout $k > 0$; $cov(y_t, y_{t-k}) = 0$.

L'avantage du model SV représenté par les équations 5 est qu'il peut être facilement transformé en un modèle linéaire, et cela en prenant le logarithme des carrés des observations, *Harvey, Ruiz et shephard (1994)*, propose de mettre $\mu = 0$, (i.e $\beta = 1$) car habituellement, la moyenne des rendement journaliers et intra-journaliers des actif est nulle. De manière à linéariser notre model d'étude, nous élevons y_t au carré et nous l'exprimons sous forme logarithmique afin de donner la représentation espace d'état de notre model d'étude, d'appliquer par la suite le filtre de Kalman tel qu'on vas le présenter dans le prochain chapitre, afin d'estimer la variabe d'état h_t qui est non observable. ■

3 Estimation du model de volatilité stochastique

3.1 Méthode du Quasi-Maximum de vraisemblance

3.1.1 Introduction

La méthode du maximum de vraisemblance n'est pas facile à implémenter dans le cas d'un modèle à volatilité stochastique, car le caractère latent de la volatilité, rend l'écriture explicite de la loi conditionnelle des observations impossible ce qui entraîne des difficultés quant à l'écriture de la fonction de vraisemblance *Ghysel (1996) et Shepard*. Une autre approche est celle de la méthode du quasi-maximum de vraisemblance (*QML*) développée par *Nelson (1988), Harvey (1994) et Ruiz (1994)*. Ils appliquent un Filtre de Kalman standard à la représentation linéaire de l'espace d'état du modèle à volatilité stochastique. Cependant cette méthode QML n'a pas une grande performance car elle ne dépend pas de la vraie vraisemblance.

3.1.2 Présentation de la méthode

D'une façon générale, l'écriture de la vraisemblance pour une série chronologique consiste à décomposer la densité jointe en produit de densités conditionnelles.

La vraisemblance s'écrit donc

$$l(y_1, \dots, y_t; \theta) = \prod_{t=2}^n f_t(y_t/y_1, \dots, y_{t-1}) f_1(y_1)$$

lorsque les lois conditionnelles et la loi de la première observation y_1 sont

gaussiennes càd :

$$\begin{aligned} f_1(y_1) &= \frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}} \exp\left(-\frac{y_1^2}{2\sigma_1^2(\theta)}\right) \\ f_t(y_t/y_1, \dots, y_{t-1}) &= \frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left(-\frac{(y_t - m_t(\theta))^2}{\sigma_t^2(\theta)2\sigma_t^2(\theta)}\right) \end{aligned}$$

où $\sigma_t^2(\theta) = \text{Var}(y_t/y_1, \dots, y_{t-1})$, $m_t(\theta) = E(y_t/y_1, \dots, y_{t-1})$ pour $2 \leq t \leq n$, et $E(y_1) = 0$, $\text{Var}(y_1) = \sigma_1^2(\theta)$

On peut donc écrire le log-vraisemblance :

$$\begin{aligned} \log[l(y_1, \dots, y_t; \theta)] &= \sum_{t=2}^n \log[f_t(y_t/y_1, \dots, y_{t-1})] + \log(f_1(y_1)) \\ &= \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}} \exp\left(-\frac{y_1^2}{2\sigma_1^2(\theta)}\right)\right) + \sum_{t=2}^n \log\left[\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left(-\frac{(y_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right)\right] \\ &= \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}}\right) + \sum_{t=2}^n \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}}\right) + \sum_{t=2}^n \left(-\frac{(y_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right) - \frac{y_1^2}{2\sigma_1^2(\theta)} \\ &= \sum_{t=1}^n \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}}\right) + \sum_{t=1}^n \left(-\frac{(y_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right) \end{aligned}$$

d'où

$$\log[l(y_1, \dots, y_t; \theta)] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log(\sigma_t^2(\theta)) + \frac{(y_t - m_t(\theta))^2}{\sigma_t^2(\theta)}\right)$$

la méthode du Quasi-Maximum de vraisemblance, permet d'estimer la log-vraisemblance, le principe général consiste d'abord à transformer le modèle d'intérêt en forme espace d'état linéaire, et par la suite utiliser le filtre de Kalman pour approximer la log-vraisemblance. Considérons le modèle 5

$$\begin{cases} y_t = B e^{h_t/2} \varepsilon_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \\ h_1 \rightsquigarrow N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right) \end{cases} \quad (9)$$

En passant au log carrée des observations on a :

$$\begin{aligned} \log(y_t^2) &= h_t + \log(\varepsilon_t^2) \\ h_{t+1} &= \mu(1 - \phi) + \phi h_t + w_t \end{aligned}$$

Où nous considérons que le paramètre $B = 1$, et on définit $Y_t = \log(y_t^2)$, $v_t = \log(\varepsilon_t^2)$. Ce qui revient donc au système d'équations suivant

$$\begin{cases} Y_t = h_t + v_t \\ h_t = \mu(1 - \phi) + \phi h_{t-1} + w_t \end{cases} \quad (10)$$

Où v_t correspond à une distribution logarithmique *Khi2*, dont l'espérance , $E(v_t) = 1.24$ et la variance, $Var(v_t) = 4.93$.Par la suite un filtre de Kalman est appliqué au model et les équations du filtre sont donnée par

Prédiction du premier pas :

$$\begin{cases} \hat{h}_{t/t-1} = \mu(1 - \phi) + \phi h_{t-1/t-1} \\ P_{t/t-1} = \phi^2 P_{t-1/t-1} + \sigma_\eta^2 \end{cases} \quad (11)$$

Mise-à-jour :

$$\begin{cases} h_{t/t} = \hat{h}_{t/t-1} + \frac{P_{t/t-1}}{F_t} \hat{e}_t \\ P_{t/t} = P_{t/t-1} - \frac{P_{t/t-1}^2}{F_t} \end{cases} \quad (12)$$

où \hat{e}_t et F_t sont respectivement l'erreur de la prédiction et sa variance, telle que

$$\begin{cases} \hat{e}_t = Y_t - \hat{h}_{t/t-1} + 1.27 \\ F_t = P_{t/t-1} + 4.93 \end{cases}$$

Si l'on connaît les valeurs initial $h_{2/1}$ et $P_{2/1}$. on peut résoudre ce filtre récursivement, d'après *Watanabe (1999)* si $|\phi| < 1$ la moyenne et la variance inconditionnelles de h_t sont données respectivement par μ et $\frac{\sigma_\eta^2}{1-\phi^2}$. ces dernières valeurs sont généralement utilisées pour $h_{2/1}$ et $P_{2/1}$. Si de plus les termes d'erreur e_t et w_t ont une distribution normale, la log vraisemblance serait

$$\log l = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log(F_t) - \frac{1}{2} \sum_{t=1}^n \frac{\hat{e}_t^2}{F_t}$$

On pourrait par la suite optimiser cette fonction de vraisemblance pour obtenir les paramètres (μ, ϕ, σ_η)

Preuve Sous les hypothèses suivantes qu'on impose sur la représentations espace d'état donnée dans 10 :

$H_1 : \{w_t\}$ est un bruit blanc gaussien de moyenne nul et de variance σ_η^2

$H_2 : \{v_t\}$ est un bruit blanc non gaussien où ; $E[v_t] = 1.24$ et $V[v_t] = 4.93$

$H_3 : l'état initial h_1 est tiré d'une loi normal de moyenne μ et de variance $\frac{\sigma_\eta^2}{1-\phi^2}$$

$H_4 : \{v_t\}, \{w_t\}$ et h_t sont mutuellement indépendants
l'espérance et la variance inconditionnelle de h_1 sont respectivement données

par :

$$h_{1/1} = \mu, P_{1/1} = \frac{\sigma_\eta^2}{1 - \phi^2}$$

et d'après H_1 et H_4 on a :

$$\hat{h}_{t/t-1} = E[h_t/I_{t-1}] = E[\mu(1 - \phi) + \phi h_t + w_t/I_{t-1}] = \mu(1 - \phi) + \phi E[h_t/I_{t-1}]$$

d'où

$$\hat{h}_{t/t-1} = \mu(1 - \phi) + \phi \hat{h}_{t/t-1}$$

La matrice variance covariance $P_{t/t-1}$ est calculée comme suite :

$$\begin{aligned} P_{t/t-1} &= E[(h_t - \hat{h}_{t/t-1})(h_t - \hat{h}_{t/t-1})'/I_{t-1}] \\ &= E[(\phi(h_t - \hat{h}_{t/t-1}) + w_t)(\phi(h_t - \hat{h}_{t/t-1}) + w_t)'/I_{t-1}] \end{aligned}$$

d'après H_4 on a l'indépendance entre $(h_t - \hat{h}_{t/t-1})$ et w_t , d'où :

$$\begin{aligned} E[w_t(\phi(h_t - \hat{h}_{t/t-1}))'/I_{t-1}] &= E[(\phi(h_t - \hat{h}_{t/t-1}))w_t'/I_{t-1}] = 0 \\ P_{t/t-1} &= \sigma_\eta^2 + \phi^2 E[(h_t - \hat{h}_{t/t-1})(h_t - \hat{h}_{t/t-1})'/I_{t-1}] \end{aligned}$$

par conséquent

$$P_{t/t-1} = \sigma_\eta^2 + \phi^2 P_{t-1/t-1}$$

une fois disponible l'observation Y_t , les quantités précédentes sont mises à jour :

$$\begin{aligned} h_{t/t} &= E[h_t/I_t] = E[h_t - \hat{h}_{t/t-1} + \hat{h}_{t/t-1}/I_t] \\ h_{t/t} &= E[h_t - \hat{h}_{t/t-1} + \hat{h}_{t/t-1}/I_{t-1}, \hat{e}_t] \end{aligned}$$

tel que $\hat{e}_t = Y_t - \hat{Y}_{t/t-1}$, est un processus d'innovation gaussien .d'où

$$h_{t/t} = E[h_t - \hat{h}_{t/t-1}/\hat{e}_t] + \hat{h}_{t/t-1}$$

et on à alors

$$h_t - h_{t/t} = h_t - \hat{h}_{t/t-1} - E[h_t - \hat{h}_{t/t-1}/\hat{e}_t]$$

par conséquent

$$P_{t/t-1} = E[(h_t - \hat{h}_{t/t-1} - E[h_t - \hat{h}_{t/t-1}/\hat{e}_t]) \times (h_t - \hat{h}_{t/t-1} - E[h_t - \hat{h}_{t/t-1}/\hat{e}_t])']$$

On sait que le vecteur $((h_t - \hat{h}_{t/t-1})', \hat{e}_t')'$ est un vecteur gaussien de moyenne nulle et de matrice variance covariance

$$\begin{pmatrix} P_{t/t-1} & P_{t/t-1} \\ P_{t/t-1} & F_t \end{pmatrix}$$

alors d'apres la proposition 1 définie dans le premier chapitre on a :

$$\begin{aligned} h_{t/t} &= \hat{h}_{t/t-1} + \frac{P_{t/t-1}}{F_t} \hat{e}_t \\ P_{t/t} &= P_{t/t-1} - \frac{P_{t/t-1}^2}{F_t} \end{aligned}$$

Où \hat{e}_t est l'erreur de la prédiction calculer comme suit

$$\begin{aligned} \hat{e}_t &= Y_t - E[Y_t/I_{t-1}] = Y_t - E[h_t + v_t/I_t] \\ \hat{e}_t &= Y_t - \hat{h}_{t/t-1} + 1.27 \end{aligned}$$

et F_t , la variance de la prédiction :

$$\begin{aligned} F_t &= \text{Var}[Y_t/I_{t-1}] = \text{Var}[h_t + v_t/I_{t-1}] \\ &= P_{t/t-1} + \text{Var}(v_t/I_{t-1}) \\ F_t &= P_{t/t-1} + 4.94 \end{aligned}$$

d'où finalement la pseudo log-vraisemblance est

$$\log[l(y_1, \dots, y_t; \theta)] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log(F_t) - \frac{1}{2} \sum_{t=1}^n \frac{\hat{e}_t^2}{F_t}$$

Une fois, l'algorithme du filtre de Kalman est déroulé et les valeurs des suites $\{h_{t/t}\}_{t=1}^n, \{\hat{h}_{t/t-1}\}_{t=2}^n, \{P_{t/t}\}_{t=1}^n, \{P_{t/t-1}\}_{t=2}^n$ sont stockées, l'estimation lissée de la dernière valeur de l'échantillon $h_{n/n}$ est la dernière entrée dans $\{h_{t/t}\}_{t=1}^n$. Alors on utilise la matrice de gain $\{J_t\}_{t=1}^n$

$$J_{t-1} = P_{t-1/t-1} \phi P_{t/t-1}^{-1}$$

comme générateur, ainsi l'estimation lissée et la matrice de variance-covariance associée sont données par :

$$\begin{cases} h_{t/n} = h_{t/t} + J_t(h_{t+1/n} - h_{t+1/t}) \\ P_{t/n} = P_{t/t} + J_t(P_{t+1/n} - P_{t+1/t})J_t' \end{cases} \quad (13)$$

Remarque 21 *On peut consulter le chapitre 1 pour voir la preuve des équations 13*

On pourrait par la suite optimiser la fonction log-vraisemblance pour obtenir les paramètres $(\mu, \phi, \sigma_\eta^2)$ de notre modèle SV. La méthode QML est facile à implémenter, mais cependant, selon *Jacquier et al (1994)*, elle n'est pas performante pour des échantillons de taille finie car elle ne se base pas

sur la vraie vraisemblance. Un autre problème de cette méthode, est que le modèle qu'on choisit doit pouvoir être transformable en forme d'espace d'état linéaire, ce qui restreint le nombre de ces modèles.

Nous allons donner trois exemples d'estimation pour les paramètres $\{\sigma_\eta^2, \mu, \phi\}$, basées sur des séries d'observations $\{Y_n\}$ simulées, où on augmente la taille des échantillons, et on change les valeurs initiales pour chaque exemple de taille différente. Le tableau suivant résume les résultats obtenus,

vraies valeurs	$n = 1000$	$n = 3000$	$n = 6000$	valeurs initiales
$\phi = 0.9$	0,9195 (0.0167)	0.8875 (0.0125)	0.8988 (0.0081)	0.5
$\mu = 0$	0,2642 (0.4008)	0.2117 (0.1692)	0.0815 0.1377	0
$\sigma_\eta^2 = 1$	1,0284 (0.1697)	1.0294 (0.1030)	1.1180 0.0756	0.9
$\phi = 0.64$	0.7232 (0.0949)	0.6383 (0.0568)	0.6209 (0.0401)	0.2
$\mu = 0$	0.1036 (0.1010)	0.0270 (0.0471)	-0.0507 (0.0401)	0.01
$\sigma_\eta^2 = 0.60$	0.4048 (0.1873)	0.5151 (0.1104)	0.6747 (0.1013)	0.1
$\phi = -0.84$	-0.8353 (0.0255)	-0.8386 (0.0143)	-0.8270 (0.0306)	0.95
$\mu = 0$	0.0418 (0.0740)	-0.0059 (0.0430)	0.0394 (0.0394)	-0.21
$\sigma_\eta^2 = 2$	1.7955 (0.2516)	2.0777 (0.1627)	2.2300 (0.1239)	0.1

Tableau 1

les résultats obtenus montrent que les valeurs estimées s'approchent des vraies valeurs et l'écart-type (les valeurs entre parenthèses), l'écart-type (les valeurs entre parenthèses) montre que l'estimateur converge, et cela en diminuant quand la taille des observations augmente.

L'exemple suivant représente une estimations de la log-volatilité basées sur des données simulées de taille $n = 1000$, ou nous prenons comme vraies paramètres $(\sigma_\eta^2 = 1, \mu = 0, \phi = 0.9)$

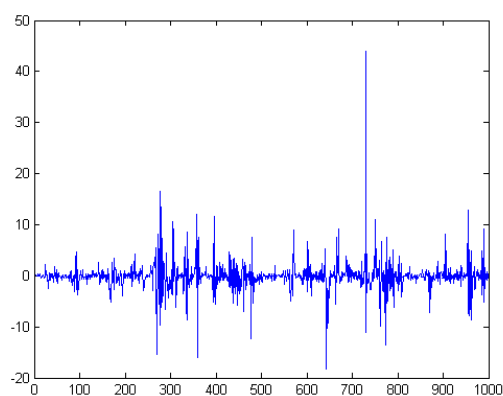


Figure 4.1.1: Simulation de y_t

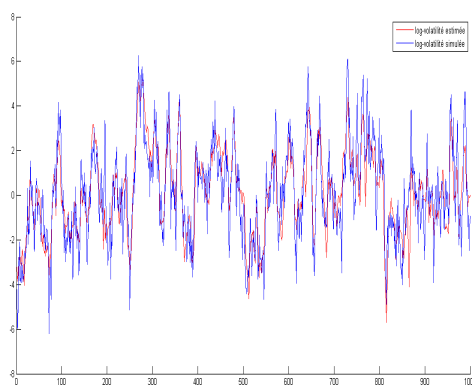


Figure 4.1.2 :Ajustement de h_t

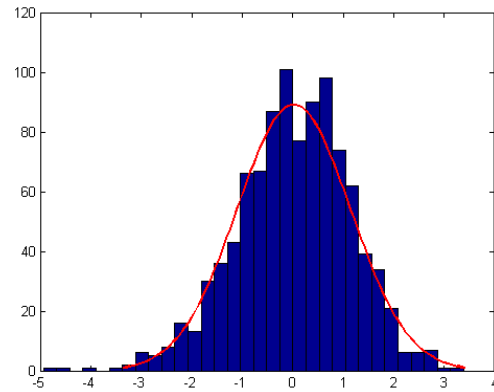


Figure 4.1.3 :Srie rediduel

Les résultats obtenues ne sont pas satisfaisantes car, d'une part la méthode du QML ne se base pas sur la vraie fonction de vraisemblance, mais une approximation, et d'autre part dans l'application du filtre de Kalman on a supposé que un bruit de distribution $\log-khi^2$, ce qui nous laisse croire que les résultats seront forcément meilleur avec un bruit gaussien.

3.2 Méthode de Monte-Carlo par Chaîne de Markov

3.2.1 Introduction

Contrairement aux autres approches classiques, adoptées dans le domaine de l'inférence statistique, à savoir le Maximum likelihood estimation (MLE) et les moindres carrés ordinaire (*MCO*), qui considèrent les paramètres comme étant vraies, l'approche bayésienne considère ces paramètres comme une variable aléatoire qui suit une loi a priori pour tenir compte de l'incertitude. Dans l'analyse bayésienne, on choisit un vecteur dont les paramètres sont des variables aléatoires qui suivent une loi a priori, on obtient, au fur et à mesure que les variables s'ajustent, une distribution a posteriori, et considérer la moyenne a postérieure pour estimer le paramètre d'intérêt. Les distributions a postérieure sont souvent difficile à calculer analytiquement, on a alors recours à des techniques numériques de Monte Carlo par Chaîne de Markov, comme l'Echantillonneur de Gibbs où l'on simule les paramètres et l'état inobservable (la volatilité) alternativement. En outre on peut rarement simuler directement par la distribution a postérieure, et par conséquent on devra utiliser l'algorithme de Métropolis Hasting. Nous rappelons dans cette section les éléments de base de l'estimation bayésienne la méthode de calculs numérique *MCMC*, ainsi que la procédure d'estimation du modèle 5.

3.2.2 Présentation de la méthode

Comme nous l'avons déjà vue dans le chapitre 1, le meilleur estimateur de Bayes associé au coût quadratique $C(\delta, y)$ et a une loi a priori π est formellement défini comme la moyenne a posteriori

$$\hat{\theta} = E[\theta|y] = \int_{\Theta} \theta \pi(\theta|y) d\theta$$

Des difficultés peuvent apparaître lors de la recherche d'une solution analytique de cette intégrale car souvent la loi a posteriori $\pi(\theta|y)$ n'est pas toujours disponible. Une solution pour calculer une intergale de la forme :

$$\int_{\Theta} g(\theta) \pi(\theta|y) d\theta$$

Où $g(\theta)$ est une fonction quelconque θ ou l'espérance existe et y le vecteur des observations

de tirer du fait que la loi à posteriori $\pi(.|y)$ est une densité de probabilité et de générer $\theta^{(m)}$ pour $m = 1, \dots, M$ suivant $\pi(\theta|y)$. Par la loi des grands nombres on a alors la convergence

$$\lim_{m \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M g(\theta^{(m)}) = \int_{\Theta} g(\theta) \pi(\theta|y) d\theta$$

lorsque la loi $\pi(\theta|y)$ n'est pas asaiment simulable par des algorithmes usuels comme la méthode d'acceptation-rejet, il existe une technique fort générale, intitulée méthode de **Metropolis-Hasting** d'après les travaux de *Metropolis et al. (1954)* et de *Hasting (1970)*, qui permet d'obtenir des réalisations approximatives de $\pi(\theta|y)$ et des estimateurs convergents de quatités a posteriori

$E[g(\theta)|y]$. Cette méthode est fondée sur une simulation itérative et markovienne de $\theta^{(m)}$ par la transition suivante :

Conditionnellement a $\theta^{(m)}$ générer : (14)

$$\theta^{(m+1)} = \begin{cases} \theta^* \sim q(\theta^*/\theta^{(m)}) \text{ avec probabilité } p \\ \theta^{(m)} \text{ sinon} \end{cases}$$

Où la loi instrumental $q(./\theta^{(m)})$, représente une densité conditionnelle arbitraire et

$$p = \min \left\{ 1, \frac{\pi(\theta^*|y)q(\theta^{(m)}|\theta^*)}{\pi(\theta^{(m)}|y)q(\theta^*|\theta^{(m)})} \right\}$$

La règle d'acceptation ou de rejet du candidat sert généralement à s'assurer que la chaîne se déplace dans la bonne direction. La validation de cette méthode découle du resultat suivant

Lemme 22 *Lorsque $\text{supp } \pi(./y) \subset \text{supp } q(./\theta)$, la chaîne de Markov $(\theta^{(m)})$ produite par 14 est irréductible ergodique et de distribution stationnaire $\pi(./y)$*

L'algorithme de **Metropolis-Hasting** appartient à la famille des méthodes de Monte Carlo par Chaîne de Markov (MCMC), qui sont fondé sur le même principe de génération d'une chaîne de Markov homogène dont la loi stationnaire et la loi d'intérêt $\pi(\theta|y)$.

Echantillonneur de Gibbs Nous allons à présent considérer une méthode MCMC qui permet de générer des variable aléatoire suivant approximativement la loi a postérieur $\pi(\theta/y)$ et qui est telle que les lois instrumentales sont presque automatiquement déduites de $\pi(\theta/y)$. Fondée sur les travaux de *Geman Geman (1984)*, *Tanner et Wong (1987)* et *Gelfand et Smith (1990)*, cette approche intitulée **échantillonnage de Gibbs**, utilise les

structure hiérarchiques du model bayésien, c'est-à-dire un model bayésien $(\pi(\theta/y), \pi(\theta))$, où la loi a priori est décomposée en distribution conditionnelles. L'hypothèse sous-jacente à l'échantillonnage de Gibbs de génération du vecteur $(\theta_1, \theta_2, \dots, \theta_m)$ est que, soit les lois conditionnelles associées à $\pi(\theta/y)$:

$$\pi_j(\theta_j/\theta_l, y, l \neq j) \quad (j = 1, \dots, m)$$

sont disponibles, soit plus généralement $\pi(\theta/y)$ peut s'écrire :

$$\pi(\theta/y) = \int \pi_1(\theta/\lambda, y) \pi_2(\lambda/y) d\lambda$$

où la loi jointe $\pi_1(\theta/\lambda, y) \pi_2(\lambda/y)$ vérifie à son tour l'hypothèse de décomposition.

Cette structure hiérarchique peut être créée de toute pièce pour la mise en oeuvre de cette technique de simulation, sans être aucunement reliée au problème inférentiel. Dans sa forme la plus simple, aussi appelée **data augmentation** par *Tanner et Wong (1987)*, la contrainte sur cette complétion est de disposer des deux distributions conditionnelles $\pi(\theta/\lambda, y)$ et $\pi(\lambda/\theta, y)$, sous forme explicite, i.e correspondant à des variables effectivement simulables. La méthode de *Tanner et Wong (1987)* est alors la suivante

Initialisation : partir d'une valeur arbitraire $\lambda^{(0)}$

Itération : pour $1 \leq i \leq k$

générer $\theta^{(i)}$ suivant $\pi(\theta/\lambda^{(i-1)}, y)$ et $\lambda^{(i)}$ suivant $\pi(\lambda/\theta^{(i)}, y)$

A l'issue de ces k tirages, on dispose d'une valeur $\theta^{(k)}$ qui, pour k assez grand et quel que soit $\lambda^{(0)}$, est approximativement tirée suivant $\pi(\theta/y)$, sous des conditions minimales de régularité. On pourra consulter *Gelfand et Smith*

(1990), Roberts et Polson (1990), Smith et Roberts (1994), Tierney (1994), ou Robert (1996) pour divers énoncés de ces conditions, on a ainsi :

Lemma 23 *si les densités conditionnelles sont presque partout positives, c'est-à-dire, $\pi(\theta/y, \lambda) > 0$ sur Θ (resp $\pi(\lambda/\theta, y) > 0$ sur Λ), les suites $(\theta^{(m)})$ et $(\lambda^{(m)})$ forment des chaînes de Markov ergodique dont les uniques loi invariantes sont respectivement $\pi(\theta/y)$ et $\pi(\lambda/y)$.*

Par conséquent, la valeur initiale $\lambda^{(0)}$ dans l'algorithme ci-dessus est bien sans importance. De plus on la convergence vers la loi stationnaire est géométrique, i.e relativement rapide, ce qui autorise à prendre k relativement petit. Une fois $\theta_1 = \theta^{(k)}$ obtenue on peut alors répéter l'algorithme avec une nouvelle valeur initiale $\lambda^{(0)}$ pour obtenir une seconde valeur θ_2 et ainsi construire un échantillon simulé $(\theta_1, \theta_2, \dots, \theta_m)$ suivant approximativement la loi $\pi(\theta/y)$ et considérer la moyenne empirique $\frac{1}{m} \sum_{i=1}^m \theta_i$ comme estimateur du paramètre θ .

3.2.3 Estimation des paramètres

L'objectif de la présente section est de décrire la partie estimation de notre modèle d'étude.

$$\left\{ \begin{array}{l} y_t = \beta e^{h_t/2} \varepsilon_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \\ h_1 \sim N(\mu, \frac{\sigma_\eta}{1-\phi^2}) \end{array} \right.$$

Où $\beta = \exp(\frac{\mu}{2})$. Pour l'estimation des paramètres (ϕ, σ_η, μ) on suivras l'algorithme suivant

Algorithme 24

Initialiser $h^{(0)}, \phi^{(0)}, \mu^{(0)}, \sigma_{\eta}^{2(0)}$

Pour $m = 1, \dots, M$

- 1) Echantillonner $h_t^{(m)}$ à partir de la loi $h_t | h_{|t}, y, \phi, \mu, \sigma_{\eta}^2$ tel que $t = 1, \dots, n$
- 2) Echantillonner $\sigma_{\eta}^{2(m)}$ à partir de la loi $\sigma_{\eta}^2 | h^{(m)}, y, \phi^{(m-1)}, \mu^{(m-1)}$
- 3) Echantillonner $\phi^{(m)}$ à partir de $\phi | h^{(m)}, \mu^{(m-1)}, \sigma_{\eta}^{2(m)}$
- 4) Echantillonner $\mu^{(m)}$ à partir de $\mu | h^{(m)}, \phi^{(m)}, \sigma_{\eta}^{2(m)}$

répéter l'algorithme K fois pour avoir les échantillons $(\phi_1^{(M)}, \phi_2^{(M)}, \dots, \phi_K^{(M)}), (\mu_1^{(M)}, \mu_2^{(M)}, \dots, \mu_K^{(M)}), (\sigma_{\eta(1)}^{2(M)}, \sigma_{\eta(2)}^{2(M)}, \dots, \sigma_{\eta(K)}^{2(M)})$ qui sont issue des lois a postérie $\pi(\phi/y), \pi(\mu/y)$ et $\pi(\sigma_{\eta}^2/y)$ respectivement, ainsi pour l'estimations des paramètres nous considérons la moyenne amprique pour chaque échantillon. Dans les prochaine lignes on expliquera comme échantillonneur chaque'un des paramètres.

Echantillonnage de h_t à partir de la loi $h | h_{|t}, y, \phi, \mu, \sigma_{\eta}^2$ La partie la plus délicate de cette algorithme est d'échantillonner h_t à partir de la loi $h_t | h_{|t}, y, \phi, \mu, \sigma_{\eta}^2$ car cette opération doit être effectuée n fois pour chaque balayage. Cependant si on note $\theta = (\phi, \mu, \sigma_{\eta}^2)$ par le théoreme de bayes et après avoir enlever les termes constant on défini la densité a postérieure marginal de h_t par :

$$f(h_t | h_{|t}, y, \theta) \propto f(h_t | h_{|t}, \theta) f(y_t | h_t, \theta) \quad t = 1, \dots, n$$

on échantillonne $h = (h_1, \dots, h_t)$ en developpant une procédure d'acceptation rejet avec courbe enveloppante apres avoir retrouver les densités a droit .en effet il est clair que $f(y_t | h_t, \theta)$ est normalement distribuer de moyenne nul et de variance $\beta^2 \exp(h_t)$, alors on monter que $f(h_t | h_{|t}, \theta)$ est aussi une densité

de loi normal et on note : $f(h_t|h_{t-1}, h_{t+1}, \theta) = f_N(h_t|h_t^*, v^2)$

$$h_t^* = \mu + \frac{\phi\{(h_{t-1} - \mu) + (h_{t+1} - \mu)\}}{(1 + \phi^2)}$$

où

$$v^2 = \frac{\sigma_\eta^2}{1 + \phi^2}$$

Preuve. Sachant h_{t-1} et θ on peut démontrer que pour tout $t = 1, \dots, n-1$

$$\begin{pmatrix} h_t \\ h_{t-1} \end{pmatrix} \sim N \left\{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \Sigma \right\}$$

tel que Σ est la matrice de variance covariance $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, on a

alors le resultat suivant :

$$f(h_t|h_{t-1}, h_{t+1}, \theta) \sim N \{u_1 + \Sigma_{12} \Sigma_{22}^{-1}, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}\}$$

où $E[h_t/h_{t-1}, h_{t+1}, \theta]$ est :

$$h_t^* = \mu + \frac{\phi\{(h_{t-1} - \mu) + (h_{t+1} - \mu)\}}{(1 + \phi^2)}$$

et $V[h_t/h_{t-1}, h_{t+1}, \theta]$ est :

$$v^2 = \frac{\sigma_\eta^2}{1 + \phi^2}$$

■

Considerons maintenant que $f(y_t|h_t, \theta) = \frac{\exp(-h_t|2)}{\sqrt{2\pi\pi}} \exp\{-\frac{1}{2}y_t^2 \exp(-h_t)\}$

alors

$$\log f(y_t|h_t, \theta) = c + \log f^*(y_t, h_t, \theta)$$

tel que

$$\log f^*(y_t, h_t, \theta) = -\frac{h_t}{2} - \frac{1}{2}y_t^2 \exp(-h_t) \text{ et } c = \log \sqrt{2\pi}$$

on note que $\exp(-h_t)$ est une fonction convexe qui est toujours supérieur ou égal à $1 + (-h_t)$ alors

$$\begin{aligned} \exp\{(h_t^* - h_t) - h_t^*\} &= \exp(h_t^* - h_t) * \exp(-h_t^*) \\ \exp\{(h_t^* - h_t) - h_t^*\} &\geq (1 + h_t^* - h_t) \exp(-h_t^*) \\ -\frac{h_t}{2} - \frac{y_t^2}{2} \exp\{(h_t^* - h_t) - h_t^*\} &\leq -\frac{h_t}{2} - \frac{y_t^2}{2} (1 + h_t^* - h_t) \exp(-h_t^*) \end{aligned}$$

d'où

$$\begin{aligned} \log f^*(y_t, h_t, \theta) &\leq -\frac{h_t}{2} - \frac{y_t^2}{2} (1 + h_t^* - h_t) \exp(-h_t^*) \\ &= g^*(y_t, h_t^*, h_t, \theta). \end{aligned}$$

comme la fonction log est monotone alors $f(h_t|h_t, \theta)f(y_t|h_t, \theta)$ qui est proportionnelle à $f(h_t|h_t, \theta)f^*(y_t, h_t, \theta)$ peut être enveloppée par $f_N(h_t|h_t^*, v^2)g^*(y_t, h_t^*, h_t, \theta)$ et on note

$$f(h_t|h_t, \theta)f^*(y_t, h_t, \theta) \leq f_N(h_t|h_t^*, v^2)g^*(y_t, h_t^*, h_t, \theta)$$

On montre que les termes à droite de l'inégalité sont proportionnel à une densité de loi normal :

$$\begin{aligned} f_N(h_t|h_t^*, v^2)g^*(y_t, h_t^*, h_t, \theta) &= \frac{1}{\sqrt{2\pi}v} \exp\left\{-\frac{1}{2v^2}(h_t - h_t^*)^2\right\} \times \\ &\quad \exp\left\{-\frac{1}{2}(h_t + y_t^2(1 + h_t^* - h_t))\exp(-h_t)\right\} \\ &= \frac{1}{\sqrt{2\pi}v} \exp\left\{-\frac{1}{2v^2}(h_t^2 - 2h_th_t^* + h_tv^2 - y_t^2 \exp(-h_t^*)h_t + h_t^* + y_t^2 v^2(1 + h_t^*)\exp(-h_t^*))\right\} \\ &= \frac{1}{\sqrt{2\pi}v} \exp\left\{-\frac{1}{2v^2}\left[h_t^2 - 2h_t(h_t^* - \frac{v^2}{2} + \frac{y_t^2 v^2 \exp(-h_t^*)}{2}) + h_t^* + y_t^2 v^2(1 + h_t^*)\exp(-h_t^*) + \right.\right. \\ &\quad \left.\left.(h_t^* - \frac{v^2}{2} + \frac{y_t^2 v^2 \exp(-h_t^*)}{2})^2 - (h_t^* - \frac{v^2}{2} + \frac{y_t^2 v^2 \exp(-h_t^*)}{2})^2\right]\right\} \end{aligned}$$

et si on tiens pas comptes des constantes alors

$$f_N(h_t|h_t^*, v^2)g^*(y_t, h_t^*, h_t, \theta) \propto f_N(h_t|\mu_t, v^2)$$

où

$$\mu_t = h_t^* + \frac{v^2}{2}[y_t^2 \exp(-h_t^*) - 1]$$

Ansı afin d'échantillonner $h = (h_1, \dots, h_t)$ par une procédure d'acceptation rejet avec la courbee enveloppante $f_N(h_t|\mu_t, v^2)$ on utilise l'algorithme suivant :

Algorithme 25

Initialiser h_t et θ

Pour $t = 1, \dots, n - 1$

- 1) *Simuler un candidat $h_t^c \sim f_N(h_t|\mu_t, v^2)z$*
- 2) *Simuler $U \sim \text{Uniforme}_{[0,1]}$ et*
$$\begin{cases} \text{Si } u \leq \frac{f^*}{g^*} \text{ alors } h_{t+1} = h_t^c \\ \text{Sinon retourner a.l'étape 1.} \end{cases}$$

Echantillonnage de σ_η^2 à partir de la loi $\sigma_\eta^2|h_{/t}, y, \phi, \mu$: D'apres l'article de **S.kim N.shepard et S. chib 1998**,etant donnée la densité a priori $p(\sigma_\eta^2) \rightsquigarrow GI(\frac{\sigma_r}{2}, \frac{s_r}{2})$ ou GI est la distribution gamma invserse. σ_η^2 est échantillonner $\text{dep}(\sigma_\eta^2|h_{/t}, y, \phi, \mu)$ tel que:

$$p(\sigma_\eta^2|h_{/t}, y, \phi, \mu) \propto p(y|h, u, \phi, \sigma_\eta^2)p(h|u, \phi, \sigma_\eta^2)p(\sigma_\eta^2)$$

puisque l'état de la volatilité h subsume toutes les informations sur les paramètres $(\mu, \sigma_\eta^2, \phi)$, l'information complète de la fonction de vraisemblance

$(y|h, u, \phi, \sigma_\eta^2)$ est une constante, et par conséquent elle peut être omise. d'où

:

$$p(\sigma_\eta^2|h_{/t}, y, \phi, \mu) \propto p(h|u, \phi, \sigma_\eta^2)p(\sigma_\eta^2)$$

Cette densité a priori suit une distribution gamma inverse et on note alors :

$$\sigma_\eta^2|h_{/t}, y, \phi, \mu \sim GI\left\{\frac{n + \sigma_r}{2}, \frac{\sigma_r + (h_1 - \mu)(1 - \phi^2) + \sum_{t=1}^{n-1} ((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2}\right\}$$

Preuve.

on sait que la densité a priori $p(\sigma_\eta^2)$ suit une loi gamma inverse, il suffit multiplier par $p(h|u, \phi, \sigma_\eta^2)$ afin de retrouver la densité a posteriori $p(\sigma_\eta^2|h_{/t}, y, \phi, \mu)$

où :

$$p(h|u, \phi, \sigma_\eta^2) = p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{t-1} p(h_{t+1}|h_1, \mu, \phi, \sigma_\eta^2)$$

$p(h_{t+1}|h_1, \mu, \phi, \sigma_\eta^2)$ suit une loi normal de moyenne $\mu + \phi(h_t - \mu)$ et de variance σ_η^2 et ainsi le produit donne :

$$\begin{aligned} p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{t-1} p(h_{t+1}|h_1, \mu, \phi, \sigma_\eta^2) &= \sqrt{\frac{1-\phi^2}{2\pi\sigma_\eta^2}} \exp\left\{\frac{(1-\phi^2)}{2\sigma_\eta^2}(h_1-\mu)^2\right\} \prod_{t=1}^{t-1} \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left\{\frac{-1}{2\sigma_\eta^2}(h_{t+1}-\mu - \phi(h_t - \mu))^2\right\} \\ &= C * \exp\left\{\frac{-(1-\phi^2)}{2\sigma_\eta^2}(h_1-\mu)^2\right\} \exp\left\{\frac{-1}{2\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1}-\mu + \phi(h_t - \mu))^2\right\} \end{aligned}$$

d'où

$$p(h|u, \phi, \sigma_\eta^2) = C * \exp\left\{\frac{-1}{2\sigma_\eta^2}[(h_1-\mu)^2(1-\phi^2) + \sum_{t=1}^{n-1} (h_{t+1}-\mu + \phi(h_t - \mu))^2]\right\} \quad (15)$$

Où $c = \sqrt{1-\phi^2} \left(\frac{1}{\sqrt{2\pi\sigma_\eta^2}}\right)^n$. En multipliant par la densité a priori on obtient :

$$\begin{aligned} p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{t-1} p(h_{t+1}|h_1, \mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2) &= \sqrt{1-\phi^2} \left(\frac{1}{\sqrt{2\pi\sigma_\eta^2}}\right)^n \exp\left\{\frac{-1}{2\sigma_\eta^2}[(h_1-\mu)^2(1-\phi^2) + \sum_{t=1}^{n-1} (h_{t+1}-\mu + \phi(h_t - \mu))^2]\right\} \\ &= \frac{s_\sigma^{\frac{\sigma_r}{2}}}{\Gamma(\frac{\sigma_r}{2})} \sqrt{1-\phi^2} \left(\frac{1}{\sqrt{2\pi\sigma_\eta^2}}\right)^n (\sigma_\eta^2)^{-1-\frac{\sigma_r}{2}} \exp\left\{-\frac{s_\sigma}{2\sigma_\eta^2} - \frac{-1}{2\sigma_\eta^2}[(h_1-\mu)^2(1-\phi^2) + \sum_{t=1}^{n-1} (h_{t+1}-\mu + \phi(h_t - \mu))^2]\right\} \\ &= \left(\frac{s_\sigma}{2}\right)^{\frac{\sigma_r}{2}} \sqrt{1-\phi^2} \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{(\sigma_\eta^2)^{-1-\frac{1}{2}(\sigma_r+n)}}{\Gamma(\frac{\sigma_r}{2})} \exp\left\{\frac{-1}{\sigma_\eta^2} \left[\frac{s_\sigma}{2} + \frac{1}{2}[(h_1-\mu)^2(1-\phi^2) + \sum_{t=1}^{n-1} (h_{t+1}-\mu + \phi(h_t - \mu))^2]\right]\right\} \end{aligned}$$

d'où

$$\sigma_\eta^2 | h_{/t}, y, \phi, \mu \sim GI \left\{ \frac{n + \sigma_r}{2}, \frac{\sigma_r + (h_1 - \mu)(1 - \phi^2) + \sum_{t=1}^{n-1} ((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2} \right\}$$

Dans notre application on prend $\sigma_r = 5$ et $s_\sigma = 0.01 \times \sigma_r$.

Echantillonnage de μ à partir de la loi $\mu | h, \phi, \sigma_\eta^2$: le paramètre μ est estimé à partir de la densité à postériori donnée par la formule suivante

$$\mu | h, \phi, \sigma_\eta^2 \sim N(\hat{\mu}, \sigma_\mu^2)$$

Tel que

$$\hat{\mu} = \sigma_\mu^2 \left\{ \frac{1 - \phi^2}{\sigma_\eta^2} h_1 + \frac{1 - \phi^2}{\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \right\}$$

et

$$\sigma_\mu^2 = \sigma_\eta^2 \left\{ (n - 1)(1 - \phi)^2 + \phi(1 - \phi^2) \right\}^{-1}$$

Preuve.

d'après l'article de *S.kim N.shepard et S. chib 1998* on suppose que la loi a priori de μ suit une distribution normal de moyenne α et de variance β^2 , or on sait que

$$p(\mu | y, h, \phi, \sigma_\eta^2) \propto p(h | \mu, \phi, \sigma_\eta^2) p(\mu)$$

alors

$$= C * \exp - \frac{1}{2} \left\{ \underbrace{-2\mu \left[\frac{\alpha}{\beta^2} + \frac{1}{\sigma_\eta^2} \{ h_1(1 - \phi^2) + (\phi - 1) \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \} \right]}_A + \frac{\mu^2 + \alpha^2}{\beta^2} + \frac{1}{\sigma_\eta^2} \{ (h_1^2 + \mu^2)(1 - \phi^2) + \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t)^2 + (n - 1)\mu^2(\phi - 1)^2 \} \right\}$$

en tenant compte que des termes multiplier par μ on aura :

$$p(\mu|y, h, \phi, \sigma_\eta^2) \propto \exp - \frac{1}{2} \left\{ \underbrace{-2\mu \left[\frac{\alpha}{\beta^2} + \frac{1}{\sigma_\eta^2} \left\{ h_1(1 - \phi^2) + (\phi - 1) \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) + \frac{\alpha}{\beta^2} \right\} \right]}_A + \underbrace{\mu^2 \left[\frac{1}{\sigma_\eta^2} (1 - \phi^2) + (n - 1)(\phi - 1)^2 + \frac{1}{\beta^2} \right]}_B \right\}$$

d'où

$$p(\mu|y, h, \phi, \sigma_\eta^2) \propto N\left(\frac{A}{B}, \frac{1}{B}\right)$$

▪

Echantillonnage de ϕ à partir de la loi $\phi|h, \mu, \sigma_\eta^2$: La procédure d'échantillonnage du paramètre ϕ à partir de la densité a postériori conditionnelle $p(\phi|h, \mu, \sigma_\eta^2)$ utilisé par *Chib et Greenberg , 1994* est basé sur l'algorithme de Metropolis-Hasting , nous allons brièvement la présenter . Soit $\phi = 2\phi^* - 1$ tel que ϕ^* suit une distribution bêta de paramètre $(\phi^{(1)}, \phi^{(2)})$ (dans notre application on prendra $(\phi^{(1)} = 20, \phi^{(2)} = 1.5)$). Par conséquent la densité a priori de ϕ est proportionnelle à :

$$p(\phi) \propto \left\{ \frac{1 + \phi}{2} \right\}^{\phi^{(1)} - 1} \left\{ \frac{1 - \phi}{2} \right\}^{\phi^{(2)} - 1} \quad \text{où} \quad \phi^{(1)}, \phi^{(2)} > \mathbb{R} \frac{1}{2}$$

d'après le théorème de Bayes la densité a postériori conditionnelle de ϕ est proportionnelle à

$$p(\phi|y, h, \mu, \sigma_\eta^2) \propto p(h|\mu, \phi, \sigma_\eta^2)p(\phi)$$

où $p(h|\mu, \phi, \sigma_\eta^2)$ est donner dans 15, la méthode consiste a proposer un can-

didat ϕ^* suivant une distribution normal de moyenne $\hat{\phi}$ et de variance V_ϕ tel que :

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1} (h_{t+1} - \mu)(h_t - \mu)}{\sum_{t=1}^{n-1} (h_t - \mu)^2} \quad \text{et} \quad V_\phi = \sigma_\eta^2 \left\{ \sum_{t=1}^{n-1} (h_t - \mu)^2 \right\}^{-1}$$

et par la suite , etant donnée la valeur actuelle $\phi^{(i-1)}$, simuler un candidat ϕ^* et l'accepter avec une probabilité $\exp \{g(\phi^*) - g(\phi^{i-1})\}$ où

$$g(\phi) = \log p(\phi) - \frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2)$$

Nous allons a présent donner trois exemples d'estimations du model SV, par la méthode MCMC basés sur des données simulée et afin d'illustrer la convergence de notre programme, on vas faire augmenter la taille des échantillons de données, le tableau suivant résume les resultats obtenue.

vraie valeurs	$n = 50$	$n = 100$	$n = 200$	$n = 500$	valeurs initiales
$\sigma_\eta^2 = 0.05$	0,0776 (0,0644)	0,0654 (0,0369)	0,0618 (0,0237)	0,0585 (0,0137)	0.1
$\mu = 0.0078$	0,0114 (0,3510)	0,0124 (0,1573)	0,0084 (0,0949)	0,0072 (0,0468)	0.0053
$\phi = 0.65$	0,7770 (0,1717)	0,7097 (0,1243)	0,6841 (0,1007)	0,6487 (0,0853)	0.75
$\sigma_\eta^2 = 0.02$	0,1063 (0,1891)	0,0667 (0,1005)	0,0387 (0,0494)	0,0216 (0,0185)	0.009
$\mu = 0$	0,0138 (0,2044)	0,0041 (0,1039)	0,0086 (0,0575)	0,0087 (0,0276)	0.01
$\phi = 0.2$	0,4867 (0,3559)	0,3696 (0,2425)	0,3012 (0,1740)	0,2132 (0,1126)	0.1
$\sigma_\eta^2 = 0.02$	0,1112 (0,1838)	0,0606 (0,0810)	0,0371 (0,0322)	0,0260 (0,0113)	0.03
$\mu = 0$	0,0113 (0,4695)	0,0160 (0,0694)	0,0119 (0,0393)	0,0135 (0,0235)	0.01
$\phi = -0.40$	-0,0336 (0,4695)	-0,1558 (0,3400)	-0,2729 (0,2156)	-0,3738 (0,1235)	-0.99

Les valeurs entre parenthèses représente la racine moyenne des erreurs au carrées (Root Mean Squart error (*RMSE*)), plus il est petit meilleur est l'estimation, on remarque qu'il diminue a chaque fois qu'on augmente la taille des observations, ce qui prouve la convergence des estimateurs vers leurs vrais valeurs respectives.

Le graphe suivant resume l'evolution du *RMSE* pour le premier exemple

$(\sigma_\eta^2 = 0.05, \mu = 0.0078, \phi = 0.65)$.

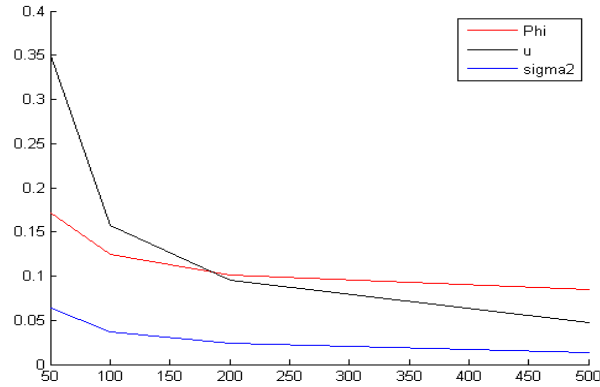


Figure 1 évolution du RMSE en fonction de la taille des observations

On remarque qu'on a une bonne estimation à partir de $n = 500$, ces résultats pourraient probablement être améliorés en augmentant la taille des observations, mais nous n'avons pas disposé de la puissance de calcul suffisante pour le faire. Les graphes suivants représentent les simulations obtenues pour chaque paramètre et les histogrammes de leurs distributions marginales respectives.

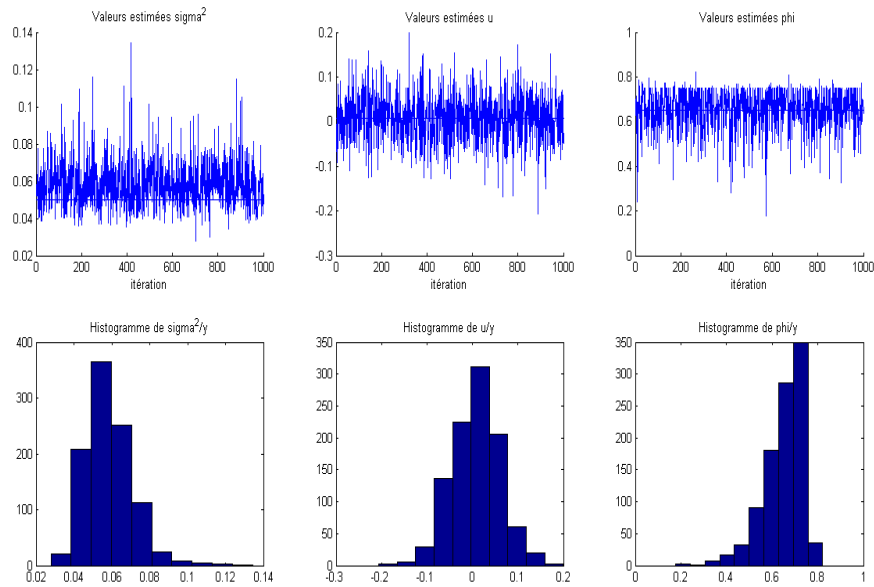


Figure 2 histogrammes des distributions marginales plus les valeurs estimées, pour chaque paramètre.