

N° d'ordre : 21/2012-M/MT

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie "Houari Boumediene"
Faculté de Mathématiques



MEMOIRE

Présenté pour l'obtention du diplôme de **MAGISTER**

En : MATHEMATIQUES

Spécialité : **Recherche Opérationnelle** : Méthodes Stochastiques

Par : **ZEGGADA Hamza**

Etude des modèles de volatilité stochastique non linéaires

Soutenu publiquement le 25/01/2012, devant le jury composé de :

BOUROUBI	Sadek	Professeur	U.S.T.H.B.	Président
AKNOUCHE	Abdelhakim	Professeur	U. S. T. H. B.	Dir. Mémoire
MIHOUBI	Miloud	Maître de Conférence /A	U. S. T. H. B.	Examineur
REBBOUH	Amar	Maître de Conférence /A	U. S. T. H. B.	Examineur

Etude des Modèles de Volatilité Stochastique Non Linéaires

ZEGGADA Hamza

Directeur de mémoire AKNOUCHE Abdelhakim
Département de Recherche Opérationnelle
Faculté de Mathématiques,
Université des Sciences et de la Technologie Houari Boumediene,
U. S. T. H. B.

Table des matières

Introduction générale	1
1 Outils mathématiques	3
1.1 Processus stochastiques	4
1.1.1 Introduction	4
1.1.2 Processus aléatoire	5
1.1.3 Equations aux récurrences stochastiques	12
1.2 Modèle Espace d'Etat et Filtre de Kalman	15
1.2.1 Introduction	15
1.2.2 Modèle espace d'état	16
1.2.3 Filtre De Kalman	25
1.3 Analyse Bayésienne et Chaînes de Markov	32
1.3.1 Introduction	32
1.3.2 Principe de l'analyse Bayésienne	32
1.3.3 Chaînes de Markov Discrètes	35
1.4 Aperçu sur les modèles de volatilité	41
1.4.1 Introduction	41
1.4.2 Modèles de moyenne conditionnelle	41
1.4.3 Processus Bruit Blanc (White Noise)	42
1.4.4 Classe des modèles ARMA	43

1.4.5	Processus autorégressif d'ordre p	44
1.4.6	Processus Moyenne Mobile d'ordre q (Moving Average)	45
1.4.7	Conclusion	47
1.5	Modèles de Variance Conditionnelle de Volatilité	48
1.5.1	Modèle ARCH (q)	48
1.5.2	Modèle GARCH (p, q)	50
2	Modèles de Volatilité Stochastique	53
2.1	Modèle canonique	53
2.1.1	Introduction	53
2.1.2	Présentation du modèle	54
2.1.3	Etude des propriétés probabilistes du modèle	57
3	Estimation du modèle de volatilité stochastique	65
3.1	Méthode du Pseudo-Maximum de Vraisemblance	65
3.1.1	Introduction	65
3.1.2	Présentation de la méthode	65
3.1.3	Estimation des paramètres	73
3.2	Méthode de Simulation Monte Carlo par Chaînes de Markov (MCMC)	76
3.2.1	Introduction	76
3.2.2	Présentation de la méthode	76
3.2.3	Estimation des paramètres	78
3.3	Application du modèle SV sur un cas réel	84
3.3.1	Introduction	84
3.3.2	Estimation par la méthode QML	85
3.3.3	Estimation par la méthode MCMC	86
3.3.4	Comparaison des QML et MCMC	87

TABLE DES MATIÈRES

iii

4 Conclusion et Perspectives

88

Bibliographie

91



Je tiens tout d'abord à remercier M^r A. AKNOUCHE, pour l'honneur qu'il m'a fait en acceptant de m'encadrer. Ses conseils précieux ont permis une bonne orientation dans la réalisation de ce modeste travail.

Je teins également à remercier M^r S. BOUROUBI d'avoir accepté de présider le jury de ce mémoire.

Je remercie M^r M. MIHOUBI et M^r A. REBBOUH d'avoir accepté de faire partie du jury et consacrer leurs temps à la lecture et à la correction de ce mémoire.

Mes remerciements les plus vifs vont tout particulièrement à mes parents.

Enfin, merci à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.



Je dédie ce modeste travail :

A mon très cher père .

A ma très chère mère.

A mon frère et ma soeur.

A tous mes amis.

H. ZEGGADA

Introduction générale

Les modèles de volatilité ont connu un grand essor ces trois dernières décennies grâce à leur aptitude à modéliser non seulement l'évolution des valeurs d'un phénomène mais aussi leur variabilité, variabilité dont le modèle adéquat est la variance conditionnelle. Une spécification essentielle de ces modèles, notamment les modèles de type ARCH (Autoregressive conditionnellement hétéroscédastiques), est que la variance conditionnelle est considérée comme fonction du passé du phénomène sous-jacent. Cependant, cette dépendance uniquement du passé présente une sérieuse limitation tant le présent n'est pas intégré dans l'évolution de la variabilité. C'est l'une des principales motivations de l'introduction, dans les années quatre vingt, d'un autre type de modèles de volatilité où la variance conditionnelle est plutôt fonction d'un processus exogène non observable évoluant selon une certaine structure dont la plus répandue est la structure markovienne homogène. Cette classe de modèles dits de volatilité stochastique, puisque la volatilité n'est pas observable au présent, est considérée comme une bonne alternative des modèles de volatilité observable, à l'instar des modèles ARCH/GARCH qui reposent sur l'adéquation entre les concepts de variance conditionnelle et de 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 et elle implique des limitations importantes sur les propriétés dynamiques.

A l'opposé, les modèles dits à volatilité stochastique font de la volatilité une variable latente (non observable) possédant une dynamique propre.

Cependant, la formulation de ces modèles, tout à fait adéquate aux théories financières, présente des difficultés dans leur mise en œuvre statistique sur la base de données réelles. La méthode classique du maximum de vraisemblance s'avère très coûteuse voire impossible

puisque l'évaluation se fait par rapport à toutes les trajectoires possibles du processus non observable de volatilité. D'autres méthodes telles les méthodes des moments simple et généralisée sont plus faciles à mettre en œuvre mais souffrent d'un déficit d'efficacité statistique comparées à la méthode du maximum de vraisemblance. Dans les années quatre vingt dix, un grand intérêt à été porté aux méthodes de simulation de Monté Carlo Markov Chain (MCMC) basées sur l'approche Bayésienne pour l'estimation et des paramètres et du processus non observable considéré comme paramètre augmenté. Ces méthodes ont prouvé leur efficacité statistique sous des délais raisonnables de mise en œuvre.

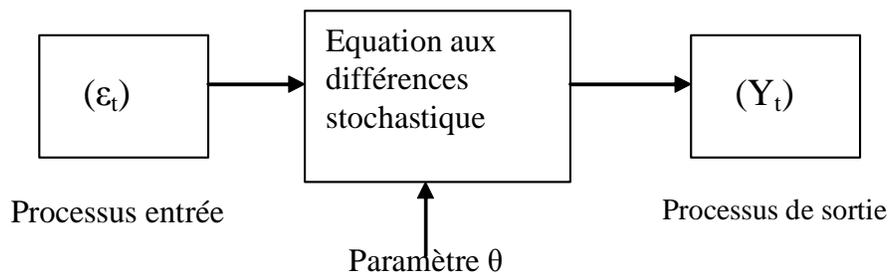
Ce mémoire porte sur l'étude probabiliste, statistique et applications des modèles de volatilité stochastique. Le modèle étudié consiste en un produit de deux processus : un processus d'innovation et le processus latent de volatilité s'exprimant via une autoregression d'ordre un. D'abord on a présenté une étude probabiliste des modèles de volatilité, à savoir les conditions d'existence de solutions avec des exigences de structure, telle la stationnarité stricte, stationnarité au second ordre, structure d'autocovariance,...

Dans une deuxième phase, on s'intéresse au problème d'estimation de tels modèles. En premier lieu on présente la méthode du quasi maximum de vraisemblance dans laquelle l'évaluation est basée sur le principe de décomposition de l'erreur de prédiction où les prédictions et les lissages nécessaires sont obtenus grâce au filtrage de Kalman. Ensuite, on étudie le problème d'estimation sous l'approche Bayésienne selon laquelle, partant de lois à priori appropriées pour les paramètres et les latents, les lois a posteriori sont obtenus itérativement via l'échantillonnage de Metropolis-Hasting, une variante connue des méthodes MCMC. Les deux méthodes sont comparées en échantillons finis à travers des simulations et dont la meilleure s'avère la méthode Bayésienne. Enfin dans la dernière phase, on propose d'appliquer la modélisation SV sur des données financières réelles où on induit la volatilité de séries réelles, le modèle étant validé.

Chapitre 1

Outils mathématiques

La quasi-totalité des modèles de séries chronologiques sont souvent exprimés par des équations aux différences stochastiques, avec un processus entrée (le plus souvent processus d'innovation et/ou processus exogène), une forme de l'équation avec ces paramètres et une sortie qui représente une solution de cette équation et qui est elle-même un processus stochastique. L'étude des modèles de séries chronologiques revient à étudier la structure des propriétés des sorties (stationnarité, ergodicité, ergodicité géométrique, comportement des queues) en fonction des propriétés des entrées (i.i.d, bruit blanc, différence de Martingale, . . .), des paramètres, et de la forme de l'équation.



1.1 Processus stochastiques

1.1.1 Introduction

La théorie des probabilités a pour but ultime la représentation mathématique des phénomènes aléatoires. En fait, on peut classer les phénomènes aléatoires en deux catégories : les phénomènes statiques et les phénomènes évolutifs (ou dynamiques). Dans le premier cas, il s'agit d'un fait qu'on admet fixe qui ne se rattache ou ne dépend d'aucun autre fait, dusse le même fait à un autre instant ou à une autre région. Le modèle mathématique de phénomènes aléatoires statiques est la variable (ou vecteur) aléatoire. Lorsqu'on veut inférer des informations concernant la distribution de probabilité des valeurs possibles d'un phénomène aléatoire statique, on peut répéter des observations identiques et indépendantes. Cependant, si l'on désire appréhender certains phénomènes comme le prix du pétrole à un instant donné, la taille d'une population à une génération ou la consommation d'électricité à une région donnée, on peut encore obtenir un ensemble d'observations prises chaque jour, chaque génération, ou chaque région, mais à ce moment là, les hypothèses d'indépendance et d'équidistributivité ne sont plus admissibles. Il est plus judicieux, ici, de considérer que les observations prises à des instants ou à des générations ou encore à des régions différentes font partie d'une même famille infinie représentant tout le phénomène à tout instant, à chaque génération ou à chaque région. Le modèle de variable aléatoire s'avère ainsi insuffisant pour décrire les phénomènes qui évoluent par rapport à un ensemble renfermant une infinité de membres, parfois non dénombrable. Ainsi, on est amené à considérer plutôt une famille de variables aléatoires, évoluant sur un domaine, et dont chacune est destinée à décrire un moment correspondant de l'évolution du phénomène. Un tel modèle est appelé processus aléatoire ou stochastique et est donc caractérisé par une structure de dépendance entre membres de la famille et une distribution de probabilité des valeurs, en raison de l'aléa qui le caractérise. La théorie des processus aléatoires a pour but de donner les règles de base pour le calcul des probabilités d'évènements se rattachant plutôt à des processus aléatoires. La statistique des processus, quant à elle, permet sur la base d'une observation d'une partie du phénomène d'induire des informations concernant la distribution de probabilité d'un ou de plusieurs membres du processus ayant généré cette observation.

Soit (Ω, A, P) un espace de probabilité, où Ω est l'espace des réalisations possibles, A est une tribu adaptée à (c'est l'ensemble qui contient les combinaisons possibles d'évènements) et P est une mesure de probabilité définie sur A .

Les processus aléatoires qui seront présentés dans la suite de ce travail sont définis sur l'espace de probabilité (Ω, A, P) et à valeurs dans un espace d'états $(\mathbb{R}, B(\mathbb{R}))$.

1.1.2 Processus aléatoire

Définition 1.1.1 *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 qui est muni d'une tribu \mathcal{E} , est une famille de variables aléatoires $(Y_t, t \in T)$ chacune définie sur (Ω, A, P) à valeur dans (E, \mathcal{E}) .*

On peut classer les processus aléatoires selon la dénombrabilité ou non de l'espace d'état E et du domaine d'évolution T . En effet, on distingue

- Les processus à temps discret et à espace d'état discret.
- Les processus à temps continu et à espace d'état discret.
- Les processus à temps discret et à espace d'état continu.
- Les processus à temps continu et à espace d'état continu.

Dans la suite on s'intéresse au cas des processus à temps discret ($T = \mathbb{Z}$ ou $T = \mathbb{N}$) et à espace d'état fini ou infini et $E = \mathbb{Z}$ soit \mathbb{R}^d ($d \in \mathbb{N}^*$) selon le cas.

Distribution de probabilité d'un processus aléatoire

Distributions fini-dimensionnelles Il est bien connu que la structure d'un processus aléatoire $(Y_t, t \in \mathbb{Z})$ défini sur un espace de probabilité (Ω, A, P) , à valeurs dans un espace d'états $(\mathbb{R}, B(\mathbb{R}))$ et de domaine d'évolution \mathbb{Z} est caractérisée par la distribution infini-dimensionnelle $P_Y(\cdot)$ défini comme suit :

$$P_Y(.) : (B(\mathbb{R}))^{\otimes \mathbb{Z}} \rightarrow [0, 1] \quad (1.1)$$

$$\begin{aligned} \mathbb{B} &= \prod_{t \in T} B_t \rightarrow P_Y(\mathbb{B}) = P(\omega : Y(\omega) \in \mathbb{B}) \\ &= P\left(\omega : \bigcap_{t \in T} Y_t(\omega) \in B_t\right), \end{aligned}$$

avec $B_t \in B(\mathbb{R})$. Comme la distribution infini-dimensionnelle est généralement difficile à manipuler, on fait recours à un outil plus simple qui est la distribution fini-dimensionnelle. L'introduction de cette distribution est justifiée par un théorème qui stipule que dans le cas à temps discret la distribution infini-dimensionnelle est uniquement déterminée par les probabilités fini-dimensionnelles et donc la structure probabiliste d'un processus à temps discret est entièrement déterminée par les distributions fini-dimensionnelles.

Définition 1.1.2 (Fonction de répartition fini-dimensionnelle)

Pour $n \in \mathbb{N}^*$ et pour tout $t_1, t_2, \dots, t_n \in \mathbb{Z}$, la fonction de répartition fini-dimensionnelle de $(Y_t, t \in \mathbb{Z})$ est la fonction de répartition de toute sous-suite finie $(Y_{i1}, Y_{i2}, \dots, Y_{in})$. Autrement dit, c'est la fonction $F_{Y_{t_1}, \dots, Y_{t_n}}(.)$ définie sur \mathbb{R}^n à valeurs dans $[0, 1]$ par :

$$F_{Y_{t_1}, \dots, Y_{t_n}}(.) : \mathbb{R}^n \rightarrow [0, 1] \quad (1.2)$$

$$(y_1, y_2, \dots, y_n) \rightarrow F_{Y_{t_1}, \dots, Y_{t_n}}(y_1, y_2, \dots, y_n) = P\left(\omega : \bigcap_{j=1}^n Y_{t_j}(\omega) \leq y_j\right).$$

A partir de la fonction de répartition on peut définir :

i) La fonction de masse fini-dimensionnelle $P_{Y_{t_1}, \dots, Y_{t_n}}(.)$ lorsque l'espace d'état \mathbb{R} est dénombrable.

ii) La fonction densité de masse fini-dimensionnelle $f_{Y_{t_1}, \dots, Y_{t_n}}(.)$ lorsque les variables aléatoires $(Y_t, t \in \mathbb{Z})$ sont absolument continues.

Caractéristiques de la distribution d'un processus aléatoire

Comme pour les variables aléatoires, la distribution infini-dimensionnelle d'un processus aléatoire $(Y_t, t \in \mathbb{Z})$ est aussi caractérisée par certaines familles particulières définies sur \mathbb{Z} à valeurs dans \mathbb{R} , à savoir : la fonction moyenne, la fonction variance, la fonction d'autocovariance.

a. Fonction moyenne : La fonction moyenne $\mu = E(Y_t)$ est une fonction de \mathbb{Z} dans \mathbb{R} qui pour tout $t \in \mathbb{Z}$ associe l'espérance mathématique du membre Y_t . Ayant comme domaine de définition l'ensemble $D_\mu = \{t \in \mathbb{Z} : E(Y_t) < \infty\}$.

b. Fonction variance : La fonction variance σ^2 est une fonction de \mathbb{Z} dans \mathbb{R} qui pour tout $t \in \mathbb{Z}$ associe la variance mathématique du membre Y_t . Ayant comme domaine de définition l'ensemble $D_{\sigma^2} = \{t \in \mathbb{Z} : E(Y_t^2) < \infty\}$.

c. Fonction d'autocovariance : La fonction d'autocovariance $\gamma(.,.)$ est une fonction de $\mathbb{Z} \times \mathbb{Z}$ dans \mathbb{R} qui pour tout couple $(t, s) \in \mathbb{Z} \times \mathbb{Z}$ associe la covariance entre les membres Y_t et Y_s . Ayant comme domaine de définition l'ensemble $D_\gamma = \{(t, s) \in \mathbb{Z} \times \mathbb{Z} : E(Y_t Y_s) < \infty\}$.

d. Fonction d'autocorrélation : On définit également la fonction d'autocorrélation $\rho(.,.)$ comme une fonction de $\mathbb{Z} \times \mathbb{Z}$ dans $[-1, 1]$ qui pour tout couple $(t, s) \in \mathbb{Z} \times \mathbb{Z}$ associe la corrélation entre les membres Y_t et Y_s . Ayant comme domaine de définition l'ensemble $D_\rho = D_\gamma$.

e. Fonction d'autocorrélation partielle : L'autocorrélation partielle δ_h de retard h mesure la corrélation entre deux observations éloignées de h périodes quand l'influence linéaire des observations à l'intérieur a été neutralisée. Donc c'est la corrélation entre Y_t et Y_{t-h} qui n'est pas expliquée par $Y_{t+1}, \dots, Y_{t-h-1}$.

f. Fonction génératrice : On appelle fonction génératrice des moments de $(Y_t, t \in \mathbb{Z})$ la fonction $M(.,.)$ définie de \mathbb{R} dans \mathbb{R} . Le domaine de définition de $M(.,.)$ est alors donné par $D_M = \{(t, z) \in T \times \mathbb{R} : E(e^{zY_t}) < \infty\}$.

Structure de dépendance associée aux processus aléatoires

a. 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 phénomène ne changent pas dans le domaine d'évolution. De tels phénomènes peuvent être représentés par lesdits processus stationnaires.

Dans la suite nous allons évoquer la définition de la stationnarité qui a été introduite par Kintchine (1931) pour le cas d'un domaine d'évolution \mathbb{Z} quelconque et qui est dite stationnarité stricte.

Définition 1.1.3 (Stationnarité stricte)

Un processus aléatoire $(Y_t, t \in \mathbb{Z})$ défini sur un espace de probabilité (Ω, A, P) à valeurs dans un espace d'état $(\mathbb{R}, B(\mathbb{R}))$ et de domaine d'évolution \mathbb{Z} est dit strictement stationnaire, si $\forall n, h \in \mathbb{Z}, \forall y_i \in \mathbb{R}$, on a

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

tel que $t_i + h \in \mathbb{Z}, i = 1, \dots, n$. Autrement dit, si les distributions fini-dimensionnelles sont invariantes par translation admissible dans le domaine d'évolution.

Le processus strictement stationnaire le plus simple est la suite indépendante et identiquement distribuée (i.i.d).

Il est important de noter que tous les moments d'un processus strictement stationnaire, lorsqu'ils existent, sont invariants dans le temps. Le processus est ainsi dit stationnaire en tous les moments ou à l'ordre l'infini (∞). Ainsi, la définition de stationnarité stricte semble contraignante puisque lorsque tous les moments existent, elle exige l'invariance de tous ces moments par rapport au temps. De plus, elle repose sur la connaissance des lois fini-dimensionnelles du processus qui ne peuvent être connues en pratique, sauf dans des cas très spéciaux. Cependant, plusieurs propriétés probabilistes essentielles des processus aléatoires peuvent être obtenues juste à partir des deux premiers moments (lorsqu'ils existent) et

pour les moments restants, la distinction est souvent négligeable. La stationnarité de ces deux premiers moments peut être suffisante pour expliquer du moins avec bonne précision, la stationnarité dans la distribution du processus. En particulier, pour les processus dits gaussiens qui sont très répandus en pratique, la stationnarité de ces deux premiers moments est équivalente à la stationnarité stricte. C'est pourquoi, on a souvent besoin d'un concept de stationnarité moins fort mais qui peut être rencontré en pratique.

Définition 1.1.4 (Stationnarité au second ordre)

Un processus aléatoire $(Y_t, t \in \mathbb{Z})$ défini sur un espace de probabilité (Ω, A, P) à valeurs dans un espace d'états $(\mathbb{R}, B(\mathbb{R}))$ et de domaine d'évolution \mathbb{Z} est dit stationnaire au second ordre (*faiblement stationnaire, covariance stationnaire*), si

i) La fonction variance $\sigma^2(t) = \text{var}(Y_t) < \infty$ existe pour tout t dans le domaine d'évolution \mathbb{Z} (et donc nécessairement la fonction moyenne $\mu(t) = E(Y_t)$ 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 que la fonction variance $\sigma^2(t)$ est constante dans le domaine d'évolution \mathbb{Z} .

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

Définition 1 Un processus ϵ_t est dit un processus bruit blanc si :

- $E(\epsilon_t) = 0$
- $E(\epsilon_t^2) = \sigma^2$
- $E(\epsilon_t \epsilon_\tau) = 0 \quad \forall t \neq \tau$

Définition 2 Bruit blanc fort est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne zéro et de variance finie.

Définition 3 Bruit blanc faible est une suite de variables aléatoires non-corrélées de moyenne zéro et de variance finie.

b. 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 donc une propriété d'irréductibilité mais associée à des processus stationnaires. Cette propriété est très importante 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 n'est autre que la variable dégénérée représentant la moyenne commune à tous les membres.

Avant de définir formellement la notion d'ergodicité, donnons d'abord le concept d'invariance d'un espace par rapport à une transformation.

Définition 1.1.5

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

i) L'application $\mathcal{T} : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ est dite opérateur de translation si

$$\mathcal{T} (\dots, Y_0, Y_1, Y_2, \dots) = (\dots, Y_1, Y_2, Y_3, \dots).$$

ii) Un sous ensemble $A \subset \mathbb{R}^{\mathbb{Z}}$ est dit \mathcal{T} -invariant si $\mathcal{T}(A) = A$. Autrement dit, le seul espace \mathcal{T} -invariant est \mathbb{R} lui-même.

Il existe deux types d'ergodicité; ergodicité faible et ergodicité forte.

Définition 1.1.6 (Ergodicité)

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

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

pour chaque ensemble A \mathcal{T} -invariant.

Autrement dit, le seul sous espace invariant par la translation \mathcal{T} à quelques ensembles de mesures nulles près est l'ensemble $\mathbb{R}^{\mathbb{Z}}$ lui-même, ce qui exprime une certaine irréductibilité.

Lorsque le processus aléatoire $(Y_t, t \in \mathbb{Z})$ est strictement stationnaire, la suite définie

par $\frac{1}{n} \sum_{t=1}^n Y_t$ converge vers une variable aléatoire, mais lorsque 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 ergodique pour processus stationnaire et ergodique

Théorème 4 *Soit $(Y_t, t \in \mathbb{Z})$ un processus ergodique strictement stationnaire avec une moyenne finie m . Alors*

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k = m \right) = 1.$$

Le théorème suivant nous donne différentes caractérisations de l'ergodicité.

Théorème de Karlin et Taylor, (1975)

Théorème 5 *(Karlin et Taylor, 1975)*

Soit $(Y_t, t \in \mathbb{Z})$ un processus stationnaire. Les conditions suivantes sont équivalentes :

- (a) $(Y_t, t \in \mathbb{Z})$ est ergodique.
- (b) Pour chaque ensemble A \mathcal{T} -invariant par translation

$$P((Y_0, Y_1, \dots) \in A) = 0 \text{ ou } 1.$$

- (c) Pour chaque ensemble A de séquences réelles (Y_0, Y_1, \dots) ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n I((Y_j, Y_{j+1}) \in A) = P((Y_0, Y_1, \dots) \in A),$$

tel que $I(A)$ est la variable aléatoire indicatrice de l'événement A .

$$I(A) = \begin{cases} 1 & \text{si } A \text{ se produit} \\ 0 & \text{si non} \end{cases}.$$

(d) Pour chaque $k = 1, 2, \dots$ et chaque ensemble A de vecteurs réels (Y_0, \dots, Y_k) ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n I((Y_j, \dots, Y_{j+k}) \in A) = P((Y_0, Y_1, \dots, Y_k) \in A).$$

(e) Pour chaque k et chaque fonction g de $k + 1$ variables,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n g(Y_j, \dots, Y_{j+k}) = E(g(Y_0, Y_1, \dots, Y_k)),$$

à condition que l'espérance existe.

(f) Pour chaque fonction ϕ de $\mathbb{R}^{\mathbb{Z}}$ dans \mathbb{R} , on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \phi(Y_j, Y_{j+1}, \dots) = E(\phi(Y_0, Y_1, \dots)),$$

sous condition que l'espérance existe.

L'ergodicité déjà définie implique le théorème ergodique dans le sens de la convergence presque sûre. On peut avoir une version du théorème plus faible dans le sens de convergence en probabilité et en moyenne quadratique.

1.1.3 Equations aux récurrences stochastiques

La quasi-totalité des modèles de séries chronologiques sont souvent exprimés par des équations aux récurrences stochastiques de la forme suivante :

$$Y_{n+1} = A_{n+1}Y_n + B_{n+1}, \quad n \in \mathbb{Z}. \quad (1.1.1)$$

Notre objectif dans cette section est de déterminer les conditions d'existence d'une unique solution stationnaire de l'équation (1.1.1), dans le cas où la séquence (A_n, B_n) est stationnaire et ergodique.

Les propriétés de stationnarité de la quasi totalité des modèles de séries chronologiques sont liées directement aux propriétés de stationnarité de la solution de l'équation (1.1.1). Il est suffisant donc d'étudier cette équation. Brandt (1986) a établi un théorème dans lequel il a donné les conditions d'existence d'une solution strictement stationnaire pour (1.1.1). Il l'a utilisé pour le cas unidimensionnel ($d = 1$), mais la démonstration est valable pour tout d .

Théorème 6 (Brandt, 1986)

Soit $\{(A_n, B_n), n \in \mathbb{Z}\}$ un processus strictement stationnaire et ergodique tel que $E(\log^+(A_0))$ et $E(\log^+(B_0))$ sont toutes les deux finies. Supposons que l'exposant de Lyapounov γ défini par

$$\gamma = \inf \left\{ E \left(\frac{1}{n+1} \log \|A_0 A_{-1}, \dots, A_{-n}\|, n \in \mathbb{N} \right) \right\},$$

est strictement négatif. Alors, pour tout $n \in \mathbb{N}$, la série

$$Y_n = \sum_{k=0}^{+\infty} A_n \times A_{n-1} \times \dots \times A_{n-k+1} \times B_{n-k},$$

converge presque sûrement, et le processus $(Y_n, n \in \mathbb{Z})$ est l'unique solution strictement stationnaire de (1.1.1).

On se place dans le cas où (A_n, B_n) est i.i.d. et on donne quelques définitions relatives à l'équation (1.1.1).

Définition 1.1.7 Un modèle autorégressif généralisé avec des coefficients i.i.d. est un modèle

$$Y_{n+1} = A_{n+1}Y_n + B_{n+1}, \quad n \in \mathbb{Z}, \tag{1.1.2}$$

avec $\{(A_n, B_n), n \in \mathbb{Z}\}$ est une séquence de variables aléatoires i.i.d. définies dans l'espace de probabilité (Ω, A, P) qui prennent leurs valeurs dans $M(d) \times \mathbb{R}^d$. La solution de cette

équation est une séquence $(Y_n, n \in \mathbb{Z})$ de \mathbb{R}^d variables aléatoires pour laquelle (1.1.2) est vérifiée.

Dans la suite nous allons définir une solution stationnaire qui ne dépend pas du futur.

Définition 1.1.8 (Solution non anticipative)

Une solution strictement stationnaire non anticipative de (1.1.2) est un processus strictement stationnaire $(Y_n, n \in \mathbb{Z})$ qui est une solution de (1.1.2), tel que pour chaque $p \in \mathbb{Z}$, Y_p est indépendant des variables aléatoires $\{(A_n, B_n), n \in \mathbb{Z}\}$.

Passons à la définition suivante qui introduit l'irréductibilité de (1.1.2).

Définition 1.1.9 Soit H un sous espace affine de \mathbb{R}^d . Il serait dit invariant sous le modèle (1.1.2) si $\{A_{0y} + B_0, y \in H\}$ est contenue dans H presque sûrement. Le modèle (1.1.2) est appelé irréductible si \mathbb{R}^d est le seul sous espace affine invariant.

Théorème 7 (Bougerol et Picard, 1992)

Considérons le modèle (1.1.2) avec des coefficients i.i.d. Supposons que ce modèle est irréductible et qu'il a une solution strictement stationnaire et non anticipative $(Y_n, n \in \mathbb{Z})$.

Donc il s'ensuit que :

- i) $A_0, A_{-1}, \dots, A_{-k}$ converge vers 0 presque sûrement quand $k \rightarrow +\infty$.
- ii) Pour tout entier n ,

$$Y_n = \sum_{k=0}^{+\infty} A_n \times A_{n-1} \times \dots \times A_{n-k+1} \times B_{n-k},$$

où la série converge presque sûrement.

- iii) Cette solution est l'unique solution strictement stationnaire de (1.1.2).

Le théorème que nous allons donner par la suite stipule que, sous la condition d'irréductibilité, s'il existe une solution strictement stationnaire non anticipative de (1.1.2) alors l'exposant de Lyapounov γ est strictement négatif.

Théorème 8 (Bougerol et Picard, 1992)

Supposons que le modèle (1.1.2) avec les coefficients i.i.d est irréductible et que $E(\log^+ \|A_0\|)$ et $E(\log^+ \|B_0\|)$ sont toutes les deux finies. Alors (1.1.2) a une solution strictement stationnaire non anticipative si et seulement si l'exposant de Lyapounov est strictement négatif.

Proposition 1.1.1 (Bougerol et Picard, 1992)

Supposons que le modèle (1.1.2) a une solution strictement stationnaire non anticipative $(Y_n, n \in \mathbb{Z})$. Soit H le sous espace affine minimal de \mathbb{R}^d tel que $P(Y_0 \in H) = 1$. Alors H est invariant sous le modèle et n'importe quel sous espace invariant de H porte une solution non anticipative strictement stationnaire.

Le corollaire suivant est une conséquence immédiate du théorème 1.1.3, théorème 1.1.5 et de la proposition 1.1.1.

Corollaire 1.1.1 (Bougerol et Picard, 1992)

Considérons le modèle (1.1.2) avec $E(\log^+ \|A_0\|)$ et $E(\log^+ \|B_0\|)$ sont finies. Supposons qu'il existe une solution non anticipative strictement stationnaire qui n'est pas portée par l'hyperplan affine. Alors les conditions suivantes sont équivalentes :

- i) L'exposant de Lyapounov est strictement négatif.
- ii) Le modèle est irréductible.
- iii) Il existe une unique solution stationnaire.

1.2 Modèle Espace d'Etat et Filtre de Kalman

1.2.1 Introduction

L'étude des systèmes physiques émettant au cours du temps des signaux déterminés par des états internes non observés, a conduit à développer une branche des sciences de l'ingénieur les modèles dits espace-état. Dans un premier temps, ces modèles sont présentés en comparaison des techniques économétriques plus usuelles pour modéliser les séries temporelles. Les méthodes d'estimation de tels modèles sont ensuite expliquées en deux temps : l'estimation des variables d'états ou (variables cachées) d'abord, puis celle des paramètres

(avec l'algorithme EM). Les modèles espace-état intègrent la distinction entre les variables observées y_t et les variables d'état (cachées) Z_t . Ils sont constitués :

— d'une ou plusieurs équation(s) de mesure décrivant la manière dont les variables observées sont générées par les variables d'état ou bien cachées et les résidus " équation d'observation " :

$$y_t = A'x_t + H'Z_t + w_t$$

— d'une ou plusieurs équation(s) d'état décrivant la manière dont les variables d'état (cachées) sont générées à partir de leur retard et de l'innovation :

$$Z_{t+1} = F \cdot Z_t + v_{t+1}$$

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.

1.2.2 Modèle espace d'état

Plusieurs modèles dynamiques peuvent être écrits sous la forme espace d'état, la valeur d'écriture d'un modèle sous cette forme peut être appréciée par la considération d'un modèle autoregressif d'ordre 1 :

$$y_{t+1} = \phi y_t + \varepsilon_{t+1} \tag{1.1}$$

avec ε_t iid $N(0, \sigma^2)$ les valeurs futures de y dépendent de (y_t, y_{t-1}, \dots) à travers seulement la valeur courante de y_t chose qui rend l'analyse du dynamique du système extrêmement simple à savoir établir des prévisions ou évaluer la fonction de vraisemblance, par exemple par récursivité à partir de (1.1) on aura :

$$y_{t+m} = \phi^m y_t + \phi^{m-1} \varepsilon_{t+1} + \dots + \phi^1 \varepsilon_{t+m-1} + \varepsilon_{t+m} \tag{1.2}$$

pour $m = 1, 2, \dots$

de laquelle la prévision optimale en m périodes est :

$$E(y_{t+m} \setminus y_t, y_{t-1}, \dots) = \phi^m y_t \quad (1.3)$$

Le vecteur d'état est une généralisation de l'équation (1.1)

$$Z_{t+1} = F \cdot Z_t + v_{t+1} \quad (1.4)$$

Le résultat (1.2) est généralisé comme suit :

$$Z_{t+m} = F^m \cdot Z_t + F^{m-1} \cdot v_{t+1} + F^{m-2} \cdot v_{t+2} + \dots + F^1 \cdot v_{t+m-1} + v_{t+m} \quad (1.5)$$

pour $m = 1, 2$

D'où la prévision optimale en m périodes est :

$$E(Z_{t+m} \setminus Z_t, Z_{t-1}, \dots) = F^m Z_t$$

les variables d'observation y_t sont en relation avec le vecteur d'état à travers l'équation d'observation du système est :

$$y_t = A' x_t + H' Z_t + w_t \quad (1.6)$$

où $\{Z_t\}$, $\{y_t\}$, $\{v_t\}$ et $\{w_t\}$ sont des processus aléatoires de dimensions $r \times 1$, $m \times 1$, $s \times 1$, et $l \times 1$ respectivement, tels que

$$\begin{cases} E(v_t) = E(w_t) = 0, \\ E(v_t v'_{t+h}) = \delta_{h,0} Q, \\ E(w_t w'_{t+h}) = \delta_{h,0} R, \\ E(v_t x'_{t-k}) = 0, \end{cases} \quad \forall t, h \in \mathbb{Z}, \forall k \geq 0,$$

où δ désigne la fonction de Kronecker. Les matrices F, H, A sont supposées, par commodité, indépendantes de t . On peut généraliser cette représentation au cas des coefficients dépendants de t . La variable $x_t \in \mathbb{R}^n$ est le vecteur des variables exogènes.

Pour déterminer la valeur future de y on a :

$$E(y_{t+m} \setminus Z_t, Z_{t-1}, \dots, y_t, y_{t-1}, \dots) = E(A' \times x_{t+m} + H' \times \zeta_{t+m} + w_{t+m} \setminus Z_t, Z_{t-1}, \dots, y_t, y_{t-1}, \dots)]$$

$$E(y_{t+m} \setminus Z_t, Z_{t-1}, \dots, y_t, y_{t-1}, \dots) = A' x_{t+m} + H' E[Z_{t+m} \setminus Z_t, Z_{t-1}, \dots, y_t, y_{t-1}, \dots)]$$

$$E(y_{t+m} \setminus Z_t, Z_{t-1}, \dots, y_t, y_{t-1}, \dots) = A' x_{t+m} + H' F^m Z_t \quad (1.7)$$

Représentation de quelques modèles linéaires sous forme d'espace d'états :

Modèles ARMA sous forme d'espace d'états Dans ce paragraphe, nous allons représenter certains modèles linéaires de séries chronologiques les plus connus sous forme d'espace d'états. Nous commençons progressivement par les modèles simples à savoir les modèles autorégressifs (AR), modèles moyennes mobiles (MA) et enfin les modèles $ARMA$, en donnant à chaque cas plusieurs représentations possibles.

Modèles Autorégressifs (AR) sous forme d'espace d'états Considérons le modèle autorégressif d'ordre p ($AR(p)$) suivant :

$$y_t - \sum_{j=1}^p \phi_j y_{t-j} = \varepsilon_t, \quad (1.8)$$

où $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc de variance σ^2 .

Le modèle (1.8) peut s'écrire comme suit

$$\begin{cases} y_t = \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t, \\ y_{t-1} = y_{t-1}, \\ \vdots \\ y_{t-p} = y_{t-p}. \end{cases} \quad (1.9)$$

Soit α_t un vecteur, de dimension p , défini par $\alpha_t(j) = y_{t-j+1}, j = 1, \dots, p$. Alors le système

(1.9) peut s'écrire sous la forme suivante

$$\alpha_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Par conséquent, on a alors

$$\begin{cases} \alpha_t = F\alpha_{t-1} + H\varepsilon_t \\ y_t = H'\alpha_t \end{cases} \quad (1.10)$$

où

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix} \text{ et } H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

les système (1.10) est une représentation en espace d'états du modèle autorégressif (1.8).

La représentation (1.10) du modèle (1.8) n'est pas unique. Il est possible de montrer qu'il y a d'autres représentations. On peut présenter une parmi d'autres qui est très utile pour l'évaluation de la fonction de vraisemblance exacte d'un modèle *AR*. Elle a été introduite par Harvey et Phillips (1979) et utilisée par Gardner, Harvey et Phillips (1980).

Considérons le modèle (1.8) et soit α_t un vecteur, de dimension p , défini par

$$\begin{cases} \alpha_t(1) = y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \\ \alpha_t(2) = \phi_2 y_{t-1} + \phi_3 y_{t-2} + \cdots + \phi_p y_{t-p+1}, \\ \vdots \\ \alpha_t(i) = \phi_i y_{t-1} + \phi_{i+1} y_{t-2} + \cdots + \phi_p y_{t-p+i-1}, \\ \vdots \\ \alpha_t(p) = \phi_p y_{t-1}. \end{cases} \quad (1.11)$$

Alors, on a

$$\begin{cases} \alpha_t(1) = y_t = \phi_1 \alpha_{t-1}(1) + \alpha_{t-1}(2) + \varepsilon_t, \\ \alpha_t(2) = \phi_2 \alpha_{t-1}(1) + \alpha_{t-1}(3), \\ \vdots \\ \alpha_t(p) = \phi_p \alpha_{t-1}(1), \end{cases}$$

ce qui est équivalent à

$$\begin{pmatrix} \alpha_t(1) \\ \alpha_t(2) \\ \vdots \\ \alpha_t(p-1) \\ \alpha_t(p) \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{p-1} & \vdots & & \ddots & 1 \\ \phi_p & 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \alpha_{t-1}(1) \\ \alpha_{t-1}(2) \\ \vdots \\ \alpha_{t-1}(p-1) \\ \alpha_{t-1}(p) \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \varepsilon_t \quad (1.12)$$

Par conséquent, on a alors

$$\begin{cases} \alpha_t = F\alpha_{t-1} + H\varepsilon_t \\ y_t = H'\alpha_t \end{cases} \quad (1.13)$$

où

$$F = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{p-1} & \vdots & & \ddots & 1 \\ \phi_p & 0 & \cdots & \cdots & 0 \end{pmatrix} \text{ et } H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Une troisième représentation a été introduite par Pearlman (1980). Dans le cas d'un modèle *ARMA*, le vecteur d'état est de dimension inférieure à ce lui de la représentation précédente, mais souvent cette représentation est privilégiée à cause de la non corrélation des bruits. Par la suite, De Jong (2000) a montré l'avantage de cette représentation suivante dans la convergence.

Considérons de nouveau le modèle (1.8), peut être écrit sous la forme suivante

$$\begin{cases} \alpha_t = F\alpha_{t-1} + G\varepsilon_{t-1} \\ y_t = H'\alpha_t + \varepsilon_t \end{cases} \quad (1.14)$$

La différence entre (1.13) et (1.14) s'explique par l'existence d'un bruit dans l'équation des observations y_t , les rendant ainsi les bruits (dans l'équation des états et l'équation des observations) corrélés. Le vecteur G dans cette représentation est égal à $(\phi_1, \phi_2, \dots, \phi_p)'$, en

le remplaçant par sa valeur dans l'équation précédente on obtient alors

$$\begin{pmatrix} \alpha_t(1) \\ \alpha_t(2) \\ \vdots \\ \alpha_t(p-1) \\ \alpha_t(p) \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{p-1} & \vdots & & \ddots & 1 \\ \phi_p & 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \alpha_{t-1}(1) \\ \alpha_{t-1}(2) \\ \vdots \\ \alpha_{t-1}(p-1) \\ \alpha_{t-1}(p) \end{pmatrix} + \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix} \varepsilon_{t-1}$$

Le développement de l'expression précédente, se présente comme suit

$$\begin{aligned} \alpha_t(1) &= \phi_1 \alpha_{t-1}(1) + \alpha_{t-1}(2) + \phi_1 \varepsilon_{t-1} \\ &= \phi_1 (\alpha_{t-1}(1) + \varepsilon_{t-1}) + \alpha_{t-1}(2) \\ &= \phi_1 y_{t-1} + \alpha_{t-1}(2) \end{aligned}$$

$$\begin{aligned} \alpha_{t-1}(2) &= \phi_2 \alpha_{t-1}(1) + \alpha_{t-1}(3) + \phi_2 \varepsilon_{t-2} \\ &= \phi_2 y_{t-2} + \alpha_{t-1}(3) \end{aligned}$$

les remplacements successifs dans l'équation d'origine, donnent

$$\begin{cases} \alpha_t(1) = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ \alpha_t(2) = \phi_2 y_{t-1} + \phi_3 y_{t-2} + \cdots + \phi_p y_{t-p+1} \\ \vdots \\ \alpha_t(p) = \phi_p y_{t-1} \end{cases}$$

d'où

$$\alpha_t(i) = \sum_{j=0}^{p-i} \phi_{i+j} y_{t-j-1}, \quad i = 1, \dots, p \tag{1.15}$$

Modèles Autorégressifs moyennes mobiles (ARMA) sous forme d'espace d'états

Pour les modèles ARMA, la représentation en espace d'états n'est pas unique, néant moins on retrouve généralement dans la littérature deux représentations, dites représentation $\max(p, q+1)$ et $\max(p, q)$. Dans ce qui suit nous allons présenter trois représentations dont deux de $\max(p, q+1)$ et une de $\max(p, q)$.

Considérons le modèle $ARMA(p, q)$ suivant

$$y_t - \sum_{j=1}^p \phi_j y_{t-j} = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1.16)$$

où $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc de variance σ^2 .

Première représentation

La première représentation est basée sur le modèle (1.10) et plus précisément c'est une généralisation du cas d'un modèle AR au cas d'un modèle $ARMA$, ceci du fait que si $\{x_t, t \in \mathbb{Z}\}$ est processus $AR(p)$ alors $\{y_t, t \in \mathbb{Z}\}$ défini par

$$y_t = x_t - \sum_{j=1}^q \theta_j x_{t-j},$$

Où $\{y_t, t \in \mathbb{Z}\}$ est un processus $ARMA(p, q)$.

Considérons l'expression suivante $\Phi(B)y_t = \Theta(B)\varepsilon_t$, équivalente au modèle (1.16), ce qui donne

$$\begin{cases} \Phi(B)x_t = \varepsilon_t, \\ y_t = \Theta(B)x_t. \end{cases} \quad (1.17)$$

Posons $r = \max(p, q + 1)$, $\phi_j = 0$ pour $j > p$ et $\theta_j = 0$ pour $j > q$. Soit α_t le r -vecteur défini par $\alpha_t = (x_t, x_{t-1}, \dots, x_{t-r+1})'$.

Comme x_t est un $AR(r)$ alors d'après (1.10) on obtient

$$\alpha_t = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{r-1} & \phi_r \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_t$$

L'utilisation du modèle (1.17) entraîne que

$$y_t = (1, -\theta_1, \dots, -\theta_q) x_t, \quad (1.19)$$

ce qui donne

$$\begin{cases} \alpha_t = F\alpha_{t-1} + G\varepsilon_t, \\ y_t = H'\alpha_t, \end{cases} \quad (1.20)$$

où F , G , et H sont identifiées à partir des modèles (1.18) et (1.19).

Comme cas particulier, considérons le modèle $MA(q)$ suivant

$$y_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (1.21)$$

la représentation (1.17) nous donne

$$\alpha_t = \begin{pmatrix} 0 & \dots & 0 & 0 \\ 1 & \dots & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & \dots & 1 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_t,$$

avec

$$y_t = (1, -\theta_1, \dots, -\theta_q) \alpha_t.$$

Deuxième représentation

La deuxième représentation en espace d'états pour les modèles $ARMA$ est équivalente à la précédente et qui correspond à une généralisation de (1.13).

Considérons le modèle (1.16) et introduisons le r -vecteur α_t défini par

$$\alpha_t(i) = \sum_{j=0}^{r-i} (\phi_{i+j} y_{t-j+1} - \theta_{i+j-1} \varepsilon_{t-j}), \quad i = 1, \dots, r \quad (1.22)$$

où $r = \max(p, q + 1)$

Ce qui donne

$$\begin{pmatrix} \alpha_t(1) \\ \alpha_t(2) \\ \vdots \\ \alpha_t(r-1) \\ \alpha_t(r) \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \phi_{r-1} & \vdots & \vdots & \ddots & 1 \\ \phi_r & 0 & \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} \alpha_{t-1}(1) \\ \alpha_{t-1}(2) \\ \vdots \\ \alpha_{t-1}(r-1) \\ \alpha_{t-1}(r) \end{pmatrix} + \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{r-2} \\ -\theta_{r-1} \end{pmatrix} \varepsilon_t.$$

Par conséquent,

$$\begin{cases} \alpha_t = F\alpha_{t-1} + G\varepsilon_t \\ y_t = H'\alpha_t \end{cases} \quad (1.23)$$

où

$$F = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & 0 \\ \vdots & & & \ddots & 0 \\ \phi_{r-1} & 0 & 0 & & 1 \\ \phi_r & 0 & 0 & & 0 \end{pmatrix}, G = \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{r-2} \\ -\theta_{r-1} \end{pmatrix} \text{ et } H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Troisième représentation

Comme on a déjà vu dans le cas des modèles AR purs, cette représentation généralise (1.14) à condition qu'il y est une composante moyenne mobile. Elle présente l'avantage de réduire la dimension du vecteur d'états à $\max(p, q)$ au lieu $\max(p, q + 1)$, mais elle est relativement peu utilisée que les précédentes en raison de la corrélation entre les bruits de l'équation des observations et l'équation des états.

Considérons le modèle (1.16) et soient $r = \max(p, q)$ et le r -vecteur α_t défini par

$$\begin{cases} \alpha_t(1) = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_r \varepsilon_{t-r} \\ \alpha_t(2) = \phi_2 y_{t-1} + \cdots + \phi_r y_{t-r+1} - \theta_2 \varepsilon_{t-1} - \cdots - \theta_r \varepsilon_{t-r+1} \\ \vdots \\ \alpha_t(r) = \phi_r y_{t-1} - \theta_r \varepsilon_{t-1} \end{cases}$$

qui peut être écrit sous la forme suivante

$$\alpha_t(i) = \sum_{j=0}^{r-i} \phi_{i+j} y_{t-j-1} - \sum_{j=1}^{r-i} \theta_{i+j} \varepsilon_{t-j}, \quad i = 1, \dots, r \quad (1.24)$$

ou sous forme récursive

$$\begin{cases} \alpha_t(1) = \phi_1 \alpha_{t-1}(1) + \alpha_{t-1}(2) + (\phi_1 - \theta_1) \varepsilon_{t-1} \\ \vdots \\ \alpha_t(r-1) = \phi_{r-1} \alpha_{t-1}(1) + \alpha_{t-1}(r) + (\phi_{r-1} - \theta_{r-1}) \varepsilon_{t-1} \\ \alpha_t(r) = \phi_r \alpha_{t-1}(1) + (\phi_r - \theta_r) \varepsilon_{t-1} \end{cases}$$

celle-ci peut s'écrire comme suit

$$\begin{cases} \alpha_t = F\alpha_{t-1} + G\varepsilon_t \\ y_t = H'\alpha_t \end{cases} \quad (1.25)$$

où

$$F = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{r-1} & \vdots & & \ddots & 1 \\ \phi_r & 0 & \cdots & \cdots & 0 \end{pmatrix}, G = \begin{pmatrix} \phi_1 - \theta_1 \\ \phi_2 - \theta_2 \\ \vdots \\ \phi_{r-1} - \theta_{r-1} \\ \phi_r - \theta_r \end{pmatrix} \text{ et } H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

1.2.3 Filtre De Kalman

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'une façon séquentielle en utilisant la technique du maximum de vraisemblance.

Un processus admettant une représentation espace-état signifie qu'il vérifie un système d'équations comme celui définit précédemment par :

a/-équation d'état :

$$Z_{t+1} = FZ_t + v_{t+1} \quad (2.1)$$

avec $E(v_{t+1}v'_{t+1}) = Q$.

b/-équation d'observation :

$$y_t = A'x_t + H'Z_t + w_t. \quad (2.2)$$

avec $E(w_t w'_t) = R$.

Le filtre peut être décrit comme un algorithme de calcul de la prévision optimale de la valeur de Z_t sur la base des informations observées à travers la date $t - 1$. En supposant que les valeurs de F, Q, A, H et R sont toutes connues. Cette prévision optimale est dérivée d'un résultat bien connu sur les vecteurs gaussiens.

Proposition :

Soient z_1 et z_2 deux vecteurs gaussiens avec :

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \rightarrow N \left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$$

alors la distribution de z_2 conditionnée sur z_1 suit $N(m, \Sigma)$ avec

$$m = u_2 + \Omega_{21}\Omega_{11}^{-1}(z_1 - u_1) \quad (2.3)$$

$$\Sigma = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \quad (2.4)$$

Donc la prévision optimale de z_2 conditionnée par z_1 est donnée par :

$$E(z_2 \setminus z_1) = u_2 + \Omega_{21}\Omega_{11}^{-1}(z_1 - u_1) \quad (2.5)$$

avec Σ caractérisant l'erreur moyenne carrée de cette prévision telle que :

$$E[(z_2 - m)(z_2 - m) \setminus z_1] = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \quad (2.6)$$

pour appliquer ce résultat, supposons que la valeur initiale du vecteur espace d'état (ξ_1) suit une loi normale et il en est de même pour v_t et w_t .

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 α_t à partir des observations de y jusqu'à la date t ;
- (iii) Lisser, c.à.d. estimer la valeur de α_t à partir des observations de y jusqu'à la date T , avec $T > t$.

Algorithme du Filtre de Kalman :

Notons d'abord :

$$\widehat{Z}_{t|t-1} = E(Z_t | Z_{t-1})$$

$$\widehat{P}_{t|t-1} = \text{var}(Z_t | Z_{t-1})$$

Première itération :

Supposons que $Z_1 \sim N(Z_{1|0}, p_{1|0})$

a) Remarque : Si les valeurs propres sont toutes dans le cercle unité donc le vecteur défini par (2.1) est stationnaire et :

$$\begin{cases} \widehat{Z}_{1|0} = 0 \quad (r \times 1) & (2.7) \\ \widehat{p}_{1|0} = E(Z_1 Z_1') \implies \text{vec}(p_{1|0}) = [I_{r^2} - (F \otimes F)]^{-1} \text{vec}(Q) & (2.8) \end{cases}$$

Où I_{r^2} est une matrice identité ($r^2 \times r^2$)

Tième itération :

On a :

$$\begin{cases} \widehat{Z}_{t+1|t} = F \widehat{Z}_{t|t} & (2.9) \\ \widehat{P}_{t+1|t} = F \widehat{P}_{t|t} F' + Q & (2.10) \end{cases}$$

avec

$$\widehat{Z}_{t|t} = \widehat{Z}_{t|t-1} + P_{t|t-1} H (H' P_{t|t-1} H + R)^{-1} (y_t - A' x_t - H' \widehat{Z}_{t|t-1}) \quad (2.11)$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} H (H' P_{t|t-1} H + R)^{-1} H' P_{t|t-1}$$

En substituant (2.9) par (2.11) et (2.10) par (2.12) on aura

$$\widehat{Z}_{t+1|t} = F \widehat{Z}_{t|t-1} + F P_{t|t-1} H (H' P_{t|t-1} H + R)^{-1} (y_t - A' x_t - H' \widehat{Z}_{t|t-1}) \quad (2.12)$$

$$P_{t+1/t} = FP_{t/t-1}F' - FP_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}H'P_{t/t-1}F' + Q$$

En récapitulation : Le filtre de kalman est un algorithme de calcul séquentiel $\{\widehat{Z}_{t+1/t}\}_{t=1}^T$

, et $\{P_{t+1/t}\}_{t=1}^T$ où :

- $\widehat{Z}_{t+1/t}$ représente la prévision optimale de Z_{t+1} basée sur les observations $(y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1)$,
- $P_{t+1/t}$ représente l'erreure moyenne carrée de cette prévision.
- e_t : est le vecteur d'innovation de moyenne nulle et de variance $\Sigma_t = HP_{t/t-1}H' + R$.

$$e_t = y_t - E(y_t \setminus x_t F_{t-1}) = y_t - H\widehat{Z}_{t/t-1} + Ax_t = H(Z_t - \widehat{Z}_{t/t-1}) + w_t$$

$$- F_{t-1} = (y'_{t-1}, y'_{t-2}, \dots, x'_{t-1}, x'_{t-2} \dots)$$

$$- K_t = P_{t/t-1}H'(HP_{t/t-1}H + R)^{-1} : \text{représente le gain de kalman.}$$

Le filtre de Kalman :

1. Prévision L'utilisation du filtre de Kalman nous permet de calculer la prévision en m périodes du vecteur d'état par

$$\widehat{Z}_{t+m/t} = E(Z_{t+m} \setminus y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1) = F^m \widehat{Z}_{t/t-1} \quad (2.14)$$

Sachant que l'erreur de cette prévision est donnée par

$$Z_{t+m} - \widehat{Z}_{t+m/t} = F^m(Z_t - \widehat{Z}_{t/t}) + F^{m-1}v_{t+1} + F^{m-2}v_{t+2} + \dots + F^1v_{t+m-1} + v_{t+m}$$

Ce qui entraîne l'erreure moyenne carrée de la prévision suivante :

$$P_{t+m/t} = E[(Z_{t+m} - \widehat{Z}_{t+m/t})(Z_{t+m} - \widehat{Z}_{t+m/t})'] = F^m P_{t/t} (F^m)' + F^{m-1} Q (F^{m-1})' + \dots + F Q F' + Q \quad (2.15)$$

Ces résultats peuvent aussi être utilisés pour décrire la prévision en m périodes du vecteur d'observation y_{t+m} sachant que x_t est déterministe.

$$\widehat{y}_{t+m|t} = E(y_{t+m} | y_t, y_{t-1}, \dots, y_1) = A'x_{t+m} + H'F^m \widehat{Z}_{t|t} \quad (2.16)$$

Où l'erreur de cette prévision est donnée par :

$$y_{t+m} - \widehat{y}_{t+m|t} = (A'x_{t+m} + H'Z_{t+m} + w_{t+m}) - (A'x_{t+m} + H'F^m \widehat{Z}_{t|t}) = H'(Z_{t+m} - \widehat{Z}_{t+m|t}) + w_{t+m}$$

Ce qui entraîne l'erreur moyenne carrée suivante :

$$E[(y_{t+m} - \widehat{y}_{t+m|t})(y_{t+m} - \widehat{y}_{t+m|t})'] = H'P_{t+m|t}H + R \quad (2.17)$$

2. Lissage Le but de l'étude du problème de lissage consiste à construire une inférence concernant la valeur de α_t basée sur tout l'ensemble d'informations recueillies \mathfrak{S}_T . Une telle inférence est appelée l'*estimation lissée* de α_t , notée :

$$\alpha_{t|T} = E(\alpha_t | \mathfrak{S}_T) \quad \forall T > t$$

La matrice de covariance de l'erreur de cette estimation lissée est notée :

$$P_{t|T} = E[(\alpha_t - \alpha_{t|T})(\alpha_t - \alpha_{t|T})']$$

Où la matrice $P_{t|T}$ représente la matrice de covariance de l'erreur de l'estimation de α_t basée sur l'ensemble d'informations \mathfrak{S}_T à l'instant T .

Supposons que la valeur réelle de α_t soit connue, à partir de la formule de mise à jour de la projection linéaire de α_t sur l'ensemble des observations déjà utilisée, permettant de calculer le nouvel estimateur de α_t formulé comme suit :

$$E(\alpha_{t-1} | \alpha_t) = \alpha_{t-1|t-1} + \{E[(\alpha_{t-1} - \alpha_{t-1|t-1})(\alpha_t - \alpha_{t|t-1})']\} \times \{E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})']\}^{-1} (\alpha_t - \alpha_{t|t-1}) \quad (1)$$

Le premier terme du produit peut s'écrire comme suit :

$$E[(\alpha_{t-1} - \alpha_{t-1|t-1})(\alpha_t - \alpha_{t|t-1})'] = E[(\alpha_{t-1} - \alpha_{t-1|t-1})(T_t \alpha_{t-1} + R_t v_t - T_t \alpha_{t-1|t-1})'] \quad (2)$$

Sachant que R_t et v_t sont non corrélés avec α_{t-1} et $\alpha_{t-1|t-1}$ respectivement, ce qui implique :

$$E[(\alpha_{t-1} - \alpha_{t-1|t-1})(\alpha_t - \alpha_{t|t-1})'] = E[(\alpha_{t-1} - \alpha_{t-1|t-1})(\alpha_{t-1} - \alpha_{t-1|t-1})'] T_t' = P_{t-1|t-1} T_t' \quad (3)$$

En remplaçant dans l'équation 1 :

$$E(\alpha_{t-1} | \alpha_t) = \alpha_{t-1|t-1} + P_{t-1|t-1} T_t' P_{t|t-1}^{-1} (\alpha_t - \alpha_{t|t-1}) \quad (4)$$

Appelons J_{t-1} la matrice de gain du problème de lissage

$$J_{t-1} = P_{t-1|t-1} T_t' P_{t|t-1}^{-1} \quad (5)$$

En remplaçant dans l'équation 2

$$E(\alpha_{t-1} | \alpha_t) = \alpha_{t-1|t-1} + J_{t-1} (\alpha_t - \alpha_{t|t-1}) \quad (5)$$

La suite des estimations lissées $\{\alpha_{t|T}\}_{t=1}^T$ est calculée comme suit :

$$\alpha_{t|T} = \alpha_{t|t} + J_t (\alpha_{t+1|T} - \alpha_{t+1|t}) \quad (6)$$

Le déroulement de l'algorithme du filtre de Kalman nous a permis de déterminer les valeurs des suites $\{\alpha_{t|t}\}_{t=1}^T$, $\{\alpha_{t|t-1}\}_{t=2}^T$, $\{P_{t|t}\}_{t=1}^T$, $\{P_{t|t-1}\}_{t=2}^T$ lesquelles sont stockées respectivement dans chaque terme de ces suites.

L'estimation lissée de la dernière valeur de l'échantillon $\alpha_{T|T}$ est représentée par la dernière entrée dans la suite $\{\alpha_{t|t}\}_{t=1}^T$.

Notons

Soit $t = T - 1$ et $\{J_t\}_{t=1}^T$ comme générateur alors l'équation 6 devient

$$\alpha_{T-1|T} = \alpha_{T-1|T-1} + J_{T-1}(\alpha_{T|T} - \alpha_{T|T-1})$$

De la même manière en posant $t = T - 2$

$$\alpha_{T-2|T} = \alpha_{T-2|t} + J_{T-2}(\alpha_{T-1|T} - \alpha_{T-1|T-2})$$

En procédant ainsi : $t = (t - 1), (t - 2), \dots$

On détermine toutes les valeurs lissées à partir de l'échantillon $\{\alpha_{t|T}\}_{t=1}^T$.

Détermination de la matrice de covariance de l'erreur associée à l'estimation de lissage.

Calculons $\alpha_t - \alpha_{t|T}$ de l'équation (6)

$$\alpha_t - \alpha_{t|T} = \alpha_t - \alpha_{t|t} - J_t(\alpha_{t+1|T} - \alpha_{t+1|t})$$

$$\alpha_t - \alpha_{t|T} + J_t\alpha_{t+1|T} = \alpha_t - \alpha_{t|t} + J_t\alpha_{t+1|t}$$

En multipliant les deux membres de l'équation par leur transposé et en prenant leur espérance

$$\begin{aligned} E[(\alpha_t - \alpha_{t|T} + J_t\alpha_{t+1|T})(\alpha_t - \alpha_{t|T} + J_t\alpha_{t+1|T})'] &= E[(\alpha_t - \alpha_{t|t} + J_t\alpha_{t+1|t})(\alpha_t - \alpha_{t|t} + J_t\alpha_{t+1|t})'] \\ E[(\alpha_t - \alpha_{t|T})(\alpha_t - \alpha_{t|T})'] + J_t E(\alpha_{t+1|T}\alpha'_{t+1|T})J'_t &= E[(\alpha_t - \alpha_{t|t})(\alpha_t - \alpha_{t|t})'] + J_t E(\alpha_{t+1|t}\alpha'_{t+1|t})J'_t \end{aligned}$$

La simplification par la non corrélation entre $(\alpha_t - \alpha_{t|T})$, $\alpha_{t+1|T}$, $(\alpha_t - \alpha_{t|t})$ et $\alpha_{t+1|t}$ donne l'équation suivante :

$$P_{t|T} = P_{t|t} + J_t \{ -E(\alpha_{t+1|T}\alpha'_{t+1|T}) + E(\alpha_{t+1|t}\alpha'_{t+1|t}) \} J'_t$$

$\{ -E(\alpha_{t+1|T}\alpha'_{t+1|T}) + E(\alpha_{t+1|t}\alpha'_{t+1|t}) \}$ s'écrit comme suit :

$$-E(\alpha_{t+1|T}\alpha'_{t+1|T}) + E(\alpha_{t+1|t}\alpha'_{t+1|t}) = \dots$$

$$= \{ E(\alpha_{t+1}\alpha'_{t+1}) - E(\alpha_{t+1|T}\alpha'_{t+1|T}) \} - \{ E(\alpha_{t+1}\alpha'_{t+1}) - E(\alpha_{t+1|t}\alpha'_{t+1|t}) \}$$

$$\begin{aligned}
&= \{E[(\alpha_{t+1} - \alpha_{t+1|T})(\alpha_{t+1} - \alpha_{t+1|T})']\} - \{E[(\alpha_{t+1} - \alpha_{t+1|t})(\alpha_{t+1} - \alpha_{t+1|t})']\} \\
&= P_{t+1|T} - P_{t+1|t}
\end{aligned}$$

En remplaçant dans l'équation précédente

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t'$$

La détermination des valeurs de l'échantillon $\{\alpha_{t|T}\}_{t=1}^T$ s'obtient en procédant de la même manière que précédemment en commençant par $t = (T - 1), (T - 2)$ et ainsi de suite.

1.3 Analyse Bayésienne et Chaînes de Markov

1.3.1 Introduction

L'analyse bayésienne devient de plus en plus populaire grâce au développement de l'outil informatique qui devient de plus en plus puissant, à l'implémentation de la Chaîne Markovienne de Monte Carlo (MCMC) et le développement de l'algorithme de Gibbs sampling et de Metropolis-Hasting. Cette analyse repose sur des concepts clés à savoir, la densité a priori, la densité a posteriori, la proportionnalité qu'on note par \propto , la précision qui est défini comme étant l'inverse de la variance. Nous présentons donc très succinctement 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 $X = (X_1, \dots, X_n)$ un n-échantillon de variables aléatoires i.i.d de densité $f(x/\theta)$ qui dépend d'un paramètre θ inconnu. La vraisemblance des observations x_1, \dots, x_n est notée $L(\underline{x}/\theta)$ où \underline{x} désigne le n-uples (x_1, \dots, x_n) .

En statistique fréquentiste, on considère θ un paramètre fixe appartenant à un espace Θ et on l'estime sur la base des échantillons observés.

Tandis que dans l'approche bayésienne, on considère θ un paramètre aléatoire et on associe à l'information tirée de l'échantillon, une information provenant d'une autre source

(avis d'analystes, d'experts,...). Cette information additionnelle sur θ est résumée par une loi de probabilité $\pi(\cdot)$ dite "loi a priori" du paramètre θ .

L'inférence bayésienne est alors fondée sur une loi dite "loi a posteriori" de θ , $\pi(\theta/X)$ qui est une fonction de vraisemblance $L(X/\theta)$ du modèle et de la loi a priori de θ , $\pi(\theta)$.

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

Distribution a priori

Le choix de la loi a priori du paramètre du modèle est crucial pour l'analyse bayésienne. Cette première touche d'une façon directe les distributions a posteriori qui sont utilisées dans l'inférence bayésienne.

Il y'a deux modes de pensées qui sont misent à l'évidence par rapport à la loi a priori. Le premier est "subjectif". Il repose sur l'information disponible sur le paramètre obtenue des opinions des chercheurs et d'experts. Cette information est exprimée par une loi de probabilité dite "distribution a priori informative".

Cependant, la théorie bayésienne peut être appliquée même dans le cas où on ne dispose pas d'informations a priori; c'est le deuxième mode de pensée qui est plutôt "objectif".

Là encore, le paramètre est considéré comme une valeur d'une variable aléatoire et on lui assigne une loi de probabilité dite "distribution a priori non informative".

Nous nous intéressons dans notre travail à la modélisation a priori informative qui est le point le plus délicat de l'analyse bayésienne. Il existe plusieurs procédés pour obtenir des lois informatives (voir Dreesbeke et al 2002). Nous donnons la description de l'un des plus intéressants des procédés qui est celui des familles naturelles conjuguées.

Familles naturelles conjuguées

Définition 9 Une famille G de lois a priori pour θ est dite conjuguée pour la vraisemblance $L(\underline{x}/\theta)$ si pour tout $\pi_\theta(\cdot) \in G$, la loi a posteriori $\pi(\theta/X)$ est un élément de G .

L'approche conjuguée reste la solution la plus standard dans le cadre informatif. Une famille naturelle conjugée 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.

Distribution a posteriori

Soit $X = (X_1, \dots, X_n)$ un n-échantillon de variables aléatoires i.i.d de densité $f(x/\theta)$. On suppose que θ est une valeur d'une variable aléatoire Θ de densité $\pi(\theta)$ et soit

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

la fonction de vraisemblance des observations.

La distribution a posteriori de θ notée $\pi(\theta/\underline{x})$ représente la distribution conditionnelle de θ sachant les observations $x = (x_1, \dots, x_n)$.

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

où encore, on peut écrire

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

où \propto signifie proportionnel à.

La distribution a posteriori $\pi(\theta/\underline{x})$ donne alors l'information dont on dispose sur θ , après observation. Elle représente un compromis entre l'information a priori (donnée par π) et l'information tirée de l'échantillon x (donnée par $L(\underline{x}/\theta)$).

1.3.3 Chaînes de Markov Discrètes

L'hypothèse de Markov est formulée en terme de probabilités conditionnelles. Un processus est dit **markovien** si les probabilités de transition ne sont fonction que de l'état le plus récent du système. Si les probabilités sont de plus indépendantes du temps, le processus est dit **homogène**. Pour un processus markovien homogène à espace d'états discret et à temps discret, nous avons affaire à une chaîne de Markov dénombrable homogène.

En d'autres termes, une chaîne de Markov est un processus stochastique possédant la propriété markovienne. Dans un tel processus, la prédiction d'un événement futur à partir de l'événement présent ne nécessite pas la connaissance du passé.

Définition et probabilités de transition en une étape d'une chaîne de Markov

Définition 10 *Un processus stochastique $(S_t, t \in \mathbb{N})$ défini sur un espace de probabilité (Ω, P) à valeurs dans un espace d'états $E = \{1, 2, \dots, K\}$ dénombrable (K peut être infini) est dit chaîne de Markov si : $(S_t, t \in \mathbb{N})$ est une chaîne de Markov, i.e pour tout $t \in \mathbb{N}$ et tout $i, j \in E$ ($0 \leq j \leq t - 1$)*

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

i.e. 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 11 (*Matrice de transition*)

Soit $(P_t, t \geq 1)$ une suite de matrices stochastiques (i.e. : ses coefficients sont positifs et la somme sur une ligne des coefficients est égale à 1). On dit que les matrices $(P_t, t \geq 1)$ sont les matrices de transtion 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 $j \in E$,

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

Distributions fini-dimensionnelles

Par la propriété de Markov, la structure de probabilité de la chaîne $(S_t, t \in \mathbb{N})$ qui est

représentée par les distributions fini-dimensionnelles

$$\{P_{t_1, t_2, \dots, t_n}(i_1, i_2, \dots, i_n), n \in \mathbb{N}^*, t_j \in \mathbb{N}, i_j \in E, 1 \leq j \leq n\},$$

où $P_{t_1, t_2, \dots, t_n}(i_1, i_2, \dots, i_n) := P(S_{t_1} = i_1, S_{t_2} = i_2, \dots, S_{t_n} = i_n)$, peut être exprimée par un moyen plus simple.

Proposition

Les distributions fini-dimensionnelles

$$\{P_{t_1, t_2, \dots, t_n}(i_1, i_2, \dots, i_n), n \in \mathbb{N}^*, t_j \in \mathbb{N}, i_j \in E, 1 \leq j \leq n\},$$

sont entièrement déterminée par :

- i) La distribution initiale $\pi_j(0) = P(S_0 = j)$, et
- ii) La matrice de probabilités de transition P .

Probabilités de transition en n étapes

On peut définir la probabilité de transition en n étapes, P_{ij}^n , d'un état i à un état j (après n étapes) :

$$P_{ij}^{(0)} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{sinon.} \end{cases}$$

$$P_{ij}^{(1)} = P_{ij}$$

$$P_{ij}^{(n)} = P(S_{t+n} = j / S_t = i).$$

Equations de Chapman-Kolmogorov non-homogènes

Les équations de Chapman-Kolmogorov, sont données par le théorème suivant.

Théorème 12 (*D. Kannan, 1979*).

Soit $(S_t, t \geq 0)$ une chaîne de Markov avec une distribution initiale $\pi_j(0)$, $j \in E$, et les probabilités de transitions en n étapes P_{ij}^n , ($i, j \in E, n \geq 0$), alors

- i) $P(S_t = i) = \sum_{j \in E} \pi_0(j) P_{ij}^n$

ii) les équations de Chapman-Kolmogorov :

$$P_{ij}^{n+m} = \sum_{z \in E} P_{iz}^m P_{zj}^n.$$

Distributions marginales

On a vu plus haut qu'à partir d'une distribution initiale $\pi_j(0) = P(S_0 = j)$ et des probabilités de transition P_{ij} , on peut caractériser toutes les distributions fini-dimensionnelles et en particulier les distributions marginales

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

avec

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

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

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

A partir de cette dernière relation on peut énoncer le résultat suivant.

Proposition

les distributions marginales sont entièrement déterminées par les distributions π et les matrices P .

Stabilité stochastique des chaînes de Markov

Quoique les probabilités conditionnelles (de transition) $(P_{ij}(t) = P(S_{t+1} = j/S_t = i))$ sont stationnaires, il n'en est pas nécessairement de même pour les probabilités non conditionnelles (marginales) $\pi_j(t) = P(S_t = j)$. A l'exception d'un choix particulier de la distribution initiale $\pi(0)$ et d'une forme souhaitable de la matrice P , les distributions $\pi(t)$, $\pi(t+1)$,

$\pi(t+2), \dots$ ne sont généralement pas égales. Cependant lorsque les probabilités de transition (P) vérifient certaines *propriétés structurelles*, même si les distributions marginales pour des rangs petits ne sont pas égales, il existe un certain t_0 assez grand, à partir duquel on a

$$\pi(t_0) \simeq \pi(t_0 + 1) \simeq \pi(t_0 + 2) \simeq \dots$$

où la dernière approximation est entendue dans le sens où existe une distribution π telle que

$$\lim_{t \rightarrow \infty} \pi(t) = \pi,$$

indépendamment du choix de la distribution initiale $\pi(0)$, où l'égalité vectorielle précédente est traduite composante par composante. Cette relation signifie qu'à partir d'un certain rang assez grand, les distributions marginales de la chaîne sont de plus en plus approximativement égales, quelle que soit la distribution initiale, ce qui traduit une certaine stabilité (stochastique). On dit alors que la chaîne est passée à un régime stationnaire, stable ou permanent. On note que dans certains cas, cette stabilité est remarquée initialement, à partir du commencement du processus.

a-Chaîne de Markov régulière (stable)

En terme de probabilités conditionnelles, la stabilité stochastique peut s'exprimer au moyen de l'égalité suivante :

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j, \quad i, j \in E,$$

signifiant encore que quelque soit l'état initial i dans lequel se trouve la chaîne, la probabilité de transition en n étapes tend vers une limite π_j (quand n tend vers l'infini) qui dépend seulement de l'état d'arrivée j . Cette formule peut se traduire sous forme matricielle.

$$\lim_{n \rightarrow \infty} P^n = \Pi, \tag{1.3.1}$$

avec Π est une matrice limite (stochastique) de la forme

$$\begin{aligned}\Pi &= \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_K \\ \pi_1 & \pi_2 & \dots & \pi_K \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_K \end{pmatrix} \\ &= \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix}.\end{aligned}$$

Si la relation (1.2.1) est vérifiée alors on dit que la chaîne est régulière. Dans certains ouvrages, une chaîne régulière est plutôt dit ergodique. Comme l'ergodicité est un nom réservé plutôt à une certaine irréductibilité plutôt qu'une stabilité (comme dans les processus stationnaires), pour éviter une confusion de notation, on préfère utiliser le mot régulière (ou stable).

b-Distributions stationnaires

Définition 13 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 alors appelé la distribution stationnaire (ou invariante). Elles sont déterminées à partir du résultat suivant.

Proposition

Si la limite $\lim_{n \rightarrow \infty} P^n$ existe et égale à Π alors les lignes de cette matrice vérifient nécessairement le système suivant

$$\begin{cases} \pi P = \pi \\ \pi \mathbf{1} = \mathbf{1}, \end{cases}$$

où $\mathbf{1}$ est le K -vecteur dans les éléments sont égaux à l'unité.

Remarque

On a vu que le problème fondamental des chaînes de Markov homogènes est celui de la stabilité (stochastique) qu'on peut résumer à travers les trois questions suivantes :

- i) (existence) Existe-il un système de distributions stationnaires ?
- ii) (Unicité) S'il existe, est-il unique ?
- iii) (Convergence) Les distributions marginales convergent-elles vers les distributions stationnaires au sens suivant : $\lim_{t \rightarrow \infty} \pi(t) = \pi$? Ces problèmes d'existence et d'unicité des distributions stationnaires ainsi que la limite des distributions marginales vers les distributions stationnaires indépendamment de la loi initiale dépendent de certaines propriétés structurelles de la chaîne qui sont : irréductibilité, apériodicité, récurrence et positivité. Ces propriétés qu'on retrouve dans la plus part des livres pédagogiques sur les chaînes de Markov peuvent constituer des conditions suffisantes pour la stabilité d'une chaîne de Markov.

Conditions suffisantes de stabilité

Les résultats suivants, bien connus dans la littérature, donnent ainsi des conditions suffisantes pour la régularité (stabilité) des chaînes de Markov.

Théorème 14 Soit $(S_n, n \in \mathbb{N})$ une chaîne de Markov définie sur un espace de probabilité (Ω, \mathcal{A}, P) à valeurs dans un espace d'état E discret. On suppose que $(S_n, n \in \mathbb{N})$ est :

- Irréductible.
- Apériodique.
- Récurrence positive.

Alors

- i) La distribution stationnaire π existe.
- ii) Elle est unique.
- iii) $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$, $\forall i \in E$.

De plus si $\pi(0) = \pi$ alors $(S_n, n \in \mathbb{N})$ est un processus strictement stationnaire, traduisant la stabilité à partir du début.

1.4 Aperçu sur les modèles de volatilité

1.4.1 Introduction

Il est bien connu que l'un des objectifs majeurs de la modélisation des séries chronologiques est la *prévision*. Jusqu'au début des années 80, on entendait par prévision, la prévision des valeurs futures d'une série sous étude à des horizons donnés. L'outil primordial le plus utilisé pour la prévision était l'espérance conditionnelle (disponible au passé), meilleure prévision au sens du critère de l'erreur quadratique moyenne. Les modèles linéaires à coefficients constants (*ARIMA*, *SARIMA*, espace d'états...) et à coefficients dépendant du temps (*ARMA* périodiques, *ARMA_t* évolutifs, *ARMA* localement stationnaires, espace d'états évolutifs...) sont qualifiés de **modèles de moyenne conditionnelle**. Cependant, dans de nombreux domaines scientifiques où le risque est de mise et où il est difficile de prévoir les valeurs futures (bruit blanc, chaos), on s'intéresse beaucoup plus (ou on se résigne) à **prévoir** plutôt les **variations futures des valeurs** (à un horizon donné) que les valeurs elles-mêmes. De telles prévisions peuvent être exploitées comme entrées de modèles stochastiques de la recherche opérationnelle, notamment les programmes mathématiques de gestion de portefeuille, de choix d'investissements...

Dans ce cas, l'outil primordial pour la prévision de la variabilité future est plutôt la variance conditionnelle (dite aussi volatilité).

1.4.2 Modèles de moyenne conditionnelle

Nous présentons dans ce qui suit quelques résultats concernant les modèles de moyenne conditionnelle constante et plus particulièrement les modèles utilisés auparavant tels que les processus Autorégressifs de moyenne mobile (*ARMA*) . Ces processus en fait sont limités car ils ne permettent pas la prise en compte des phénomènes de variabilité de la volatilité en fonction du temps et des mécanismes d'asymétrie.

Mais juste avant on a besoin de rappeler quelques définitions et résultats concernant ces modèles classiques de la forme suivante :

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, t \in \mathbb{Z}, \quad (4.2.1)$$

avec $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un processus bruit blanc (d'innovation) de moyenne nulle et de variance finie c'est-à-dire $(Z_t \rightarrow WN(0, \sigma_\varepsilon^2))$. Pour des raisons de normalisation nous exigeons également que $\psi_0 = 1$. On dit que les représentations de la forme (4.2.1) sont causals si et seulement si $\psi_j = 0$ pour $j < 0$.

Pour t fixé les séries dans la représentation (4.2.1) convergent asymptotiquement à condition que les coefficients à valeurs réelles ψ_j satisfont la condition suivante :

$$var(Y_t) = \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty. \quad (4.2.2)$$

1.4.3 Processus Bruit Blanc (White Noise)

Le plus simple processus stationnaire en analyse des séries temporelles est appelé : processus bruit blanc $\{\varepsilon_t, t \in \mathbb{Z}\}$ qui est une suite de variables aléatoires non corrélées de moyenne nulle et de variance constante σ_ε^2 .

Le fait que les variables aléatoires $(\varepsilon_t)_{t \in \mathbb{Z}}$ soient mutuellement non corrélées (hypothèse d'orthogonalité), nous permet de donner la fonction d'autocovariance de ce processus par :

$$\gamma_\varepsilon(h) = cov(\varepsilon_t, \varepsilon_{t+h}) = E(\varepsilon_t \varepsilon_{t+h}) = \begin{cases} \sigma_\varepsilon^2 & si \ h = 0 \\ 0 & si \ nom \end{cases} .$$

Par conséquent, la fonction d'autocorrélation est donnée par :

$$\rho_\varepsilon(h) = \begin{cases} 1 & si \ h = 0 \\ 0 & si \ h \neq 0 \end{cases} .$$

1.4.4 Classe des modèles ARMA

Définition 4.4.1 Les processus linéaires les plus populaires sont des processus $ARMA(p, q)$ qui sont données par :

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, t \in \mathbb{Z}. \quad (4.4.1)$$

Comme on peut exprimer le modèle $ARMA(p, q)$ par l'équation aux différences :

$$\Phi(B) Y_t = \Theta(B) \varepsilon_t,$$

avec ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 et :

$$\begin{aligned} \Theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \text{ et } \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta_j &\in \mathbb{R}, \forall j = 1, \dots, q \text{ et } \phi_i \in \mathbb{R}, \forall i = 1, \dots, p. \end{aligned}$$

Théorème 4.4.1 (condition de stationnarité et d'inversibilité) :

1. Une condition nécessaire et suffisante pour que le processus autorégressif moyenne mobile d'ordre (p, q) soit stationnaire est que les racines de l'équation caractéristique suivante : $\Phi(Z) = 0$ soient à l'extérieur du cercle unitaire.

2. Une condition nécessaire et suffisante pour que le processus autorégressif moyenne mobile d'ordre (p, q) soit inversible est que les racine de l'équation caractéristique suivante : $\Theta(Z) = 0$ soient à l'extérieur du cercle unitaire.

Caractéristique d'un Processus ARMA (p, q) :

Les corrélogrammes simple et partiel sont un mélange des fonctions exponentielles et sinusoïdales amorties. Cependant l'identification des paramètres p et q à partir de l'étude des fonctions d'autocorrélations empiriques s'avère plus délicate.

Remarques :

1. Les processus AR , MA et $ARMA$ ne sont représentatifs que de séries stationnaires

en tendance et corrigées des variations saisonnières. De plus ces modèles ne tiennent pas compte d'une éventuelle variable exogène.

2. dans un contexte ultérieur, nous allons exposer une extension de la classe des modèles *ARMA*.

3. L'ordre (p, q) sont généralement déterminé par l'intermédiaire d'un critère de sélection d'ordre. Les paramètres ϕ_i et θ_i satisfont certaines conditions afin de garantir que l'équation (4.4.1) a une solution ce qui peut être exprimé sous la forme (4.2.1).

Parmi les cas spéciaux de processus *ARMA* (p, q) on a le processus moyenne mobile d'ordre q (noté par *MA* (q)) et le processus autorégressif d'ordre p (noté par *AR* (p)).

1.4.5 Processus autorégressif d'ordre p

Définition 4.5.1 Le processus $\{Y_t, t \in \mathbb{Z}\}$ satisfait à une représentation *AR* d'ordre p , noté *AR* (p) , s'il est solution de l'équation aux différences stochastique suivante :

$$\varepsilon_t = Y_t - \sum_{j=1}^p \phi_j Y_{t-j},$$

ou encore

$$\varepsilon_t = \Phi(B) Y_t \text{ avec } \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \text{ et } \phi_p \in \mathbb{R}^*,$$

où $\Phi(B)$ représente le polynôme de retard et ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 .

Remarque

Le degré p est appelé le degré d'autocorrélation, il indique la profondeur de la mémoire.

Théorème 4.5.1 (condition de stationnarité)

Une condition nécessaire et suffisante pour que le processus autorégressif soit stationnaire du second ordre est que les racines de l'équation caractéristique suivante $\Phi(Z) = 0$ soient à l'extérieur du cercle unitaire.

Cas particulier :

Soit le processus stationnaire $AR(1)$: Ce processus est dit processus de Markov car l'observation dépend seulement de l'observation précédente Y_{t-1} et qui s'écrit comme suit :

$$\begin{aligned}
 Y_t &= \varepsilon_t + \phi_1 Y_{t-1} & (3.4.1) \\
 &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 Y_{t-2} \\
 &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots \\
 &= \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}
 \end{aligned}$$

la condition $|\phi_1| < 1$ est évidemment nécessaire afin de justifier la dernière égalité, dans ce cas on dit (4.2.2) converge, et (4.5.1) converge asymptotiquement.

1.4.6 Processus Moyenne Mobile d'ordre q (Moving Average)

Définition 3.5.1 Le processus $\{Y_t, t \in \mathbb{Z}\}$ satisfait à une représentation moyenne mobile d'ordre q , noté : $MA(q)$, s'il est solution de l'équation aux différences stochastique suivante :

$$Y_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}.$$

En introduisant le polynôme de retard nous obtenons :

$$Y_t = \Theta(B) \varepsilon_t \text{ où } \Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \text{ et } \theta_j \in \mathbb{R}^*,$$

et ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 .

Remarque

Le modèle moyenne mobile d'ordre q ($MA(q)$), explique la valeur de la série à l'instant t par une moyenne pondérée d'aléas ε_t , jusqu'à la $q^{\text{ième}}$ période qui sont supposées être générées par un processus de type bruit blanc.

Théorème 4.6.1 (condition d'inversibilité) :

Une condition nécessaire et suffisante pour que le processus moyenne mobile $MA(q)$ soit inversible est que les racines de l'équation caractéristique suivante : $\Theta(Z) = 0$ soient à l'extérieur du cercle unitaire.

Exemple :

Soit le processus $MA(1) : Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ dans ce cas $\Theta(Z) = 1 - \theta Z$ alors : $\Theta(Z) = 0 \Rightarrow 1 - \theta Z = 0 \Rightarrow Z = \frac{1}{\theta} \Rightarrow |Z| > 1 \Rightarrow |\theta| < 1$.

Caractéristiques d'un Processus MA(q)

Un processus moyenne mobile d'ordre q est toujours stationnaire, car il est une combinaison linéaire finie d'un processus stationnaire $\{\varepsilon_t, t \in \mathbb{Z}\}$.

Pour le corrélogramme simple seuls ses q premiers termes sont différents de zéro. La fonction d'autocorrélation est dite tronquée au-delà du $q^{\text{ème}}$ retard puisque la fonction d'autocorrélation est définie par :

$$\rho_k = \begin{cases} \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^q \theta_i^2} & \text{pour } k = 0, \dots, q \\ 0 & \text{pour } k > q \end{cases} .$$

Le corrélogramme partiel est caractérisé par une décroissance exponentielle de ses termes.

Remarques

1. Un processus autorégressif est toujours inversible.
2. Il y a une équivalence entre le processus $MA(1)$ et le processus $AR(p)$ avec p infini.

En effet, dans la modélisation ARMA, l'espérance conditionnelle varie au cours du temps alors que la variance conditionnelle ne change pas.

Par exemple, soit un modèle $AR(1)$ où $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t \tag{1}$$

La moyenne et la variance inconditionnelles de Y_t s'écrivent :

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1} \quad (2)$$

$$Var(Y_t) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \quad (3)$$

Aussi :

$$E(Y_t/Y_{t-1}) = \phi_0 + \phi_1 Y_{t-1} \quad (4)$$

La moyenne conditionnelle de Y_t dépend de l'information disponible au temps $t - 1$. et n'est pas nécessairement constante. Par contre, la variance conditionnelle est fixe et ne dépend de l'information disponible au temps $t - 1$ en raison de l'hypothèse de volatilité constante :

$$Var(Y_t/Y_{t-1}) = E[(Y_t - E(Y_t/Y_{t-1}))^2/Y_{t-1}] = E[\epsilon_t^2/Y_{t-1}] = E(\epsilon_t^2) = \sigma_\epsilon^2 \quad (5)$$

1.4.7 Conclusion

Tenant compte de l'hypothèse sur les résidus soient des bruits blancs forts nous amène à ce résultat. Un bruit blanc fort implique que les résidus ont une moyenne nulle et qu'ils sont non corrélés dans le temps. De plus, tout comme la variance inconditionnelle, la variance conditionnelle est constante. Cette dernière condition est peu réaliste parce que la variabilité dans le temps des variances est un fait stylisé bien établi en finance. Ce qui nous amène à conclure que les modèles de moyenne conditionnelle (*SARIMA*, *ARMA* périodiques, *ARMA_t* évolutifs, modèles espace d'état) sont incapables de prévoir les variations (ou du moins les prévisions sont naïves, voire inutiles) puisque la variance conditionnelle, prédicteur de la variabilité, ne dépend pas du passé, seul instrument permettant d'appréhender l'avenir.

1.5 Modèles de Variance Conditionnelle de Volatilité

1.5.1 Modèle ARCH (q)

Introduction

Face aux lacunes des représentations ARMA(p,q) pour les problèmes monétaires et financiers, depuis le début des années 80, un grand intérêt a été porté aux modèles de variance conditionnelle (dits aussi modèles de volatilité) qui permettent de modéliser et prévoir la variabilité sur la base du passé des valeurs. On distingue :

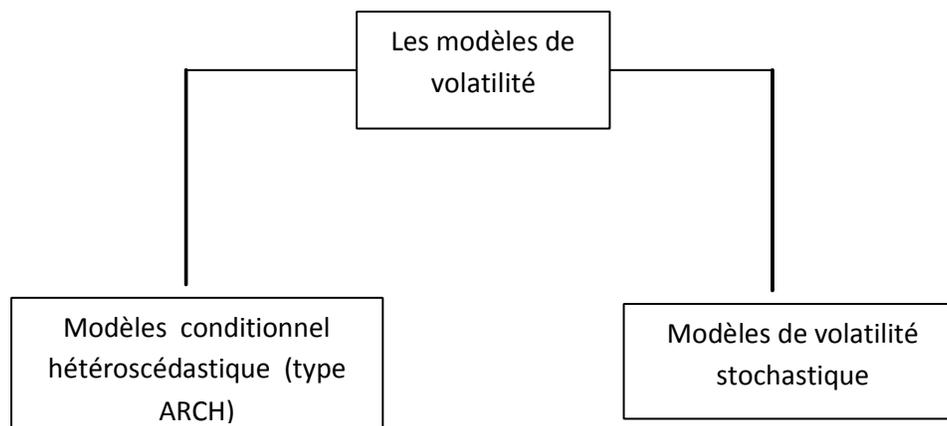
Modèles *AR* à coefficients aléatoires.

Modèles bilinéaires.

Modèles de mélange couvrant : modèles à seuil (threshold), modèles à changement de régime Markovien, mélange de modèles *AR*.

Modèles de volatilité stochastique.

Modèles de type *ARCH* : *ARCH*, *GARCH*.



L'approche ARCH/GARCH a été proposée pour prendre en compte des variances conditionnelles dépendant du temps. Le principe général consiste donc à remettre en cause la propriété d'homoscédasticité que l'on retient généralement dans le cadre du modèle linéaire.

La spécification hétéroscédastique conditionnelle ou ARCH a été initiée par Engle (1982) pour caractériser la dynamique des seconds moments conditionnels que l'on retrouve dans

la plupart des séries financières. Elle a été par la suite généralisée par Bollerslev (1986) avec ce qu'on a appelé l'hétéroscédastique conditionnelle autorégressive généralisée ou GARCH c'est le modèle le plus populaire lorsqu'il s'agit d'estimer les variances conditionnelles.

Les modèles GARCH ne se contentent pas seulement d'estimer des variances qui évoluent dans le temps, mais incorporent également les caractéristiques observés sur les séries financières (leptokurtisme, clusters de volatilité, ...).

D'autres généralisations ont été proposées par plusieurs chercheurs, on peut citer les modèles AGARCH, EGARCH, FIGARCH, GJR-GARCH, TARCH, ...

Présentation du modèle ARCH

Les économistes utilisent fréquemment des modèles estimés à l'aide des séries temporelles où la variabilité des résidus est relativement faible pendant un certain nombre des périodes successives, puis beaucoup plus grande pour un certain nombre d'autres périodes et ainsi de suite, et ce généralement sans aucune raison apparente, ce phénomène est particulièrement fréquent et visible avec les séries boursières, des taux des changes étrangers, ou d'autre prix déterminés, sur les marchés financiers, ou la volatilité semble généralement varier dans le temps. Récemment, d'importants approfondissements ont vu le jour dans la littérature pour modéliser ce phénomène. L'article novateur d'Engle (1992), expose pour la première fois le concept d'hétéroscédasticité conditionnelle autorégressive, ou ARCH. L'idée fondamentale est que la variance de l'aléa au temps t dépend de l'importance des aléas au carré des périodes passés, cependant il existe plusieurs façons de modéliser cette idée de base, la littérature correspondante est foisonnante.

Définition 5.1.1 :

Un processus $\{\epsilon_t, t \in \mathcal{I}\}$ satisfait une représentation ARCH(q) si :

$$\begin{cases} \epsilon_t = \eta_t \sqrt{h_t} \\ E(\epsilon_t^2 | \epsilon_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2, \quad \alpha_q \neq 0, \text{Où } \alpha_0 > 0, \alpha_i \geq 0 \forall i = 1, \dots, q \end{cases}$$

et $(\eta_t)_t$ désigne un bruit blanc faible tel que $E(\eta_t) = 0$ et $E(\eta_t^2) = \sigma^2$.

Pour ce type de processus on retrouve une propriété essentielle à savoir la propriété de variance conditionnelle variable dans le temps puisque :

$$Var(\epsilon_t | I_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2$$

Où I_{t-1} est la tribu engendrée par le passé du processus jusqu'au temps $t-1$.

1.5.2 Modèle GARCH (p, q)

Introduction

Plusieurs variantes du modèle ARCH ont été proposées. Une variante particulièrement utile est le modèle ARCH généralisé ou GARCH suggéré par Bollerslev (1986). Contrairement au modèle ARCH, la variance conditionnelle h_t dépend aussi bien de ses propres valeurs passées que des valeurs retardées de ϵ_t^2 .

Dans la pratique un modèle GARCH avec très peu de paramètres ajuste souvent aussi bien qu'un modèle ARCH ayant de nombreux paramètres, en particulier, un modèle simple qui fonctionne souvent très bien est le modèle GARCH (1, 1).

On considère un modèle linéaire autorégressif exprimé sous la forme suivante :

$$X_t = E(X_t | X_{t-1}) + \epsilon_t \quad \text{Où } \{\epsilon_t, t \in \mathcal{I}\} \text{ est un bruit blanc faible tel que } E(\epsilon_t) = 0 \text{ et } E(\epsilon_t \epsilon_s) = 0 \text{ si } t \neq s,$$

satisfaisant la condition de différence de martingale. $E(\epsilon_t | I_{t-1}) = 0$.

On cherche à modéliser la volatilité conditionnelle du processus de bruit $\{\epsilon_t, t \in \mathcal{I}\}$ pour tenir compte de la dynamique observée, on peut être amené à imposer une valeur élevée du paramètre q dans la modélisation ARCH (q) ce qui peut poser des problèmes d'estimation.

Il s'agit d'une difficulté semblable à celle que l'on rencontre dans les modélisations de l'espérance conditionnelle : si le théorème de Wold assure que toute série stationnaire possède une représentation de type MA, il est possible que pour une série donnée, l'ordre de cet MA soit particulièrement élevé, voir infini. Dans ce cas Box et Jenkins proposent de regagner en parcimonie en utilisant une représentation de type AR (p) ou ARMA (p, q). Pour la variance conditionnelle, Bollerslev (1986) définit ainsi le processus GARCH (p, q).

Définition 5.2.1 :

Un processus $\{\epsilon_t, t \in \mathcal{I}\}$ satisfait une représentation GARCH(p,q) si :

$$\begin{cases} \epsilon_t = \eta_t \sqrt{h_t} \\ h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \end{cases} \quad \text{Où } (\eta_t) \text{ est un bruit blanc}$$

avec les conditions $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0 \forall i = 1, \dots, q \forall j = 1, \dots, p$ suffisante pour garantir la positivité de h_t . Ainsi l'erreur du processus $\{X_t, t \in \mathcal{I}\}$ définie par le processus GARCH(p, q) admet pour moments conditionnels : $E(\epsilon_t | \epsilon_{t-1}) = 0, Var(\epsilon_t | I_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$.

Tout comme le modèle ARCH, on peut exprimer le processus ϵ_t^2 sous la forme d'un processus ARMA défini dans une innovation.

$$u_t = \epsilon_t^2 - h_t$$

En introduisant cette notation dans l'équation d'un GARCH (p, q), il dévient :

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

D'où l'on tire que :

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

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

Remarque :

Le degré de p apparaît comme un degré de moyenne mobile de la représentation ARMA dans ϵ_t^2 . A partir de cette représentation, on peut calculer de façon aussi simple les moments et les moments conditionnels du processus d'erreur $\{\epsilon_t, t \in \mathcal{I}\}$ mais aussi du processus $\{X_t, t \in \mathcal{I}\}$.

Exemple :

Considérons le modèle $GARCH(1, 1)$.

$$\begin{cases} y_t = -6.37 \times 10^{-5} + \epsilon_t \\ \sigma_t^2 = 9.97 \times 10^{-7} + 0.81\sigma_{t-1}^2 + 0.15\epsilon_{t-1}^2 \end{cases}$$

Ce modèle est le résultat de l'estimation de la série des observations quotidiennes du taux de change Deutschmark / Livre sterling par la modélisation $GARCH(1, 1)$. On génère un échantillon d'observations de taille 1000, pour estimer les paramètres du modèle (la série d'innovation). Ces paramètres sont simulés et estimés à l'aide des deux fonctions `garchsim` et `garchfit` définies sur Matlab. On obtient comme résultat une comparaison entre les innovations, les variances conditionnelles et les rendements du modèle.

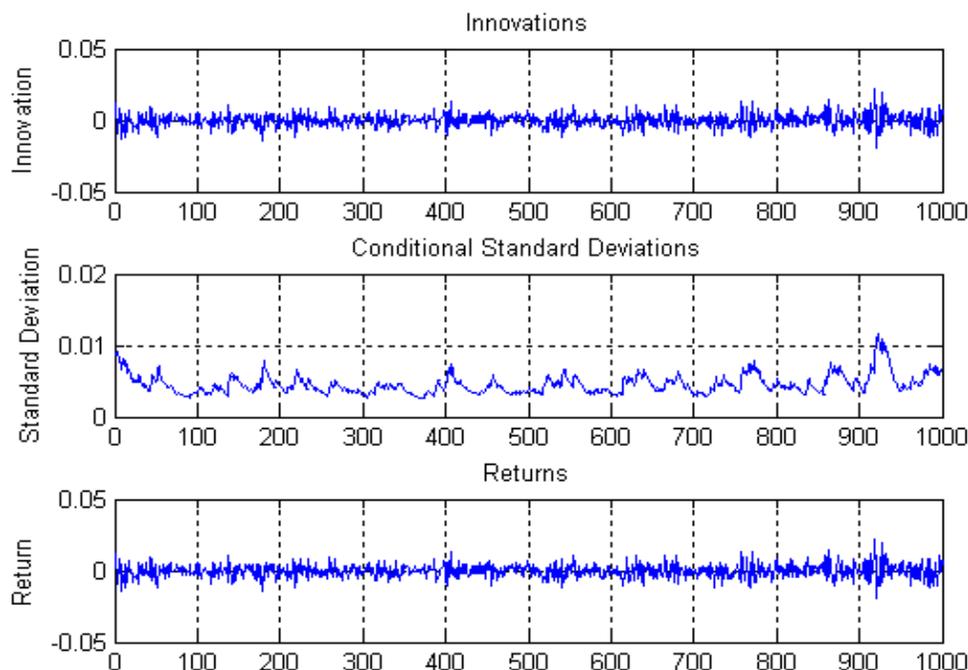


Figure 1.5.1 : Estimation du modèle $GARCH(1, 1)$, comparaison entre les innovations, variances conditionnelles et les rendements.

Chapitre 2

Modèles de Variance Conditionnelle Non Observable

2.1 Modèle canonique

2.1.1 Introduction

Nous avons vu que les modèles GARCH reposent sur l'adéquation entre les concepts de variance conditionnelle et de 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 et elle implique des limitations importantes sur les propriétés dynamiques.

A l'opposé, les modèles dits à volatilité stochastique font de la volatilité une variable latente (non observable) possédant une dynamique propre. Le processus observé, ϵ_t , et sa volatilité, σ_t , sont liés par l'équation

$$\epsilon_t = \sigma_t \eta_t$$

Où (η_t) est un bruit blanc indépendant, généralement supposé indépendant du processus (σ_t) .

On complète le modèle en spécifiant la dynamique de σ_t , a priori quelconque pourvu qu'elle soit compatible avec la positivité de cette variable. Comme dans le cas GARCH, il existe une spécification simple, couramment utilisée, suffisamment riche pour reproduire les

principales caractéristiques des séries financières et susceptible d'extensions. Cette spécification, appelée modèle canonique, consiste à supposer que le logarithme de la volatilité, ou de son carré, suit un modèle AR(1) avec terme constant. Dans ces modèles, les processus (ε_t) et (σ_t) ne sont évidemment pas indépendants mais, contrairement au cas GARCH, la variable $h_t = \sigma_t^2$ ne s'interprète plus comme la variance de ε_t conditionnelle à son passé : celle-ci n'a pas, dans ces modèles, une forme explicite. Il existe cependant de nombreuses analogies entre les deux classes de processus et celles-ci doivent être vues comme concurrentes, en particulier pour la modélisation des séries financières.

2.1.2 Présentation du modèle

La première apparition du modèle à volatilité stochastique canonique dans la littérature économétrique éte par Taylor en 1986. Cependant, les méthodes statistiques permettant de l'utiliser en pratique n'ont été développées que dans le courant des années 90. Nous commençons par passer en revue les propriétés élémentaires du modèle canonique.

Considérons le modèle

$$\begin{cases} \varepsilon_t = \sqrt{h_t} \eta_t \\ \log h_t = \omega + \beta \log h_{t-1} + \sigma v_t \end{cases} \quad (1)$$

Où $|\beta| < 1$, (η_t) et (v_t) sont deux suites indépendantes de variables iid centrées et de variances égales à 1.

Le modèle à volatilité stochastique canonique, a été le plus étudié dans la littérature en raison de sa simplicité et de ses liens avec des modèles en temps continu. Notons que l'écriture en logarithme de la dynamique de la volatilité assure la positivité de h_t sans qu'il soit nécessaire de contraindre les coefficients ω, β et σ . On pourra cependant supposer $\sigma > 0$ sans perte de généralité.

La volatilité h_t étant dépendante de son passé, il est clair que le module de ε_t dépend également de celui de ses valeurs passées.

L'interprétation des paramètres dans ce modèle se présente comme suit :

Le paramètre β joue le rôle du coefficient de persistance. En effet, il est facile de voir, que lorsque β est proche de 1, un choc positif sur la volatilité (assimilable à une grande valeur positive de v_t) a généralement pour effet de maintenir la volatilité à un niveau élevé sur plusieurs périodes. L'effet du choc est évidemment atténué, puisque $\beta < 1$, et tend à disparaître.

Un choc négatif a inversement un impact négatif sur la volatilité. Si β est proche de 0, l'effet du choc est transitoire, la volatilité dépendant peu de ses valeurs passées. Enfin, si β 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 β comme peu de réalistes pour les applications.

L'interprétation des autres coefficients est plus immédiate.

Le paramètre ω est un facteur d'échelle pour la volatilité.

Le paramètre σ mesure l'amplitude des innovations de h_t , il peut s'interpréter comme la volatilité de la log-volatilité.

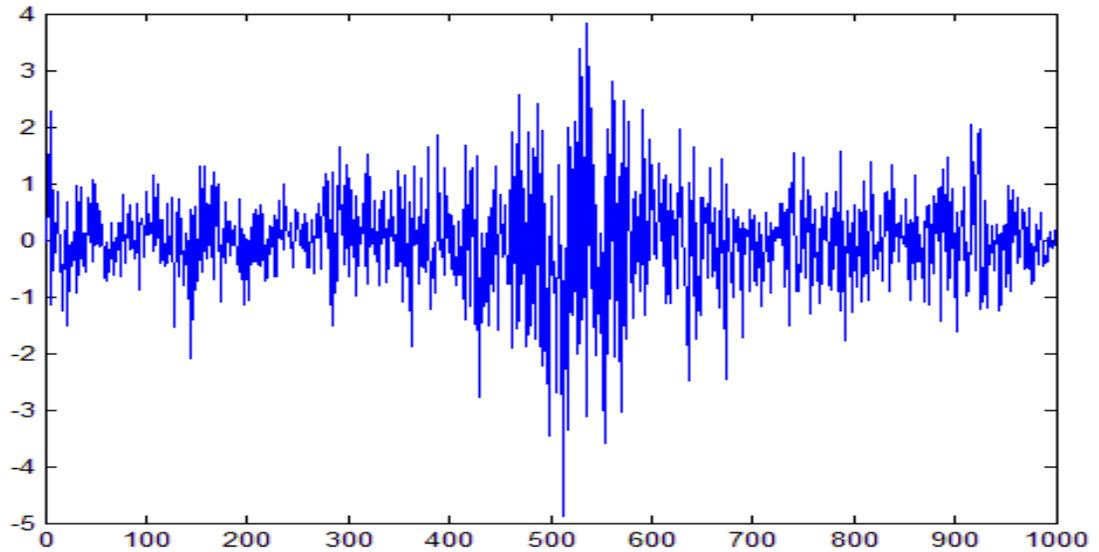


Figure 2.1.1 : Simulation d'une série de taille 1000 du modèle SV de paramètres $\beta = 0.999$
 $\sigma = 0.01$

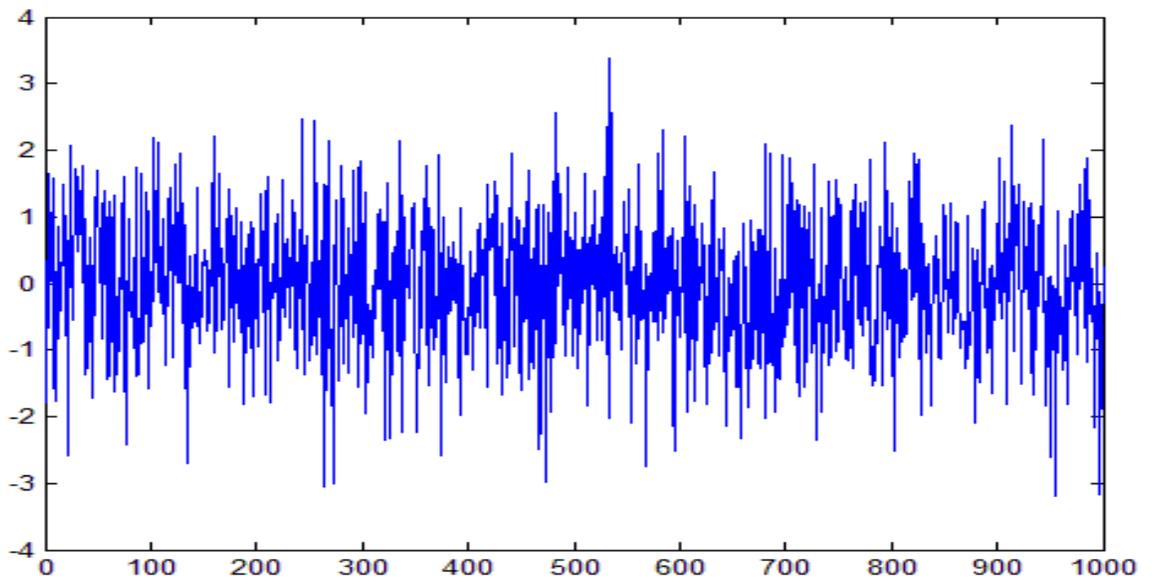


Figure 2.1.2 : Simulation d'une série de taille 1000 du modèle SV de paramètre $\beta = 0.0001$
 $\sigma = 0.9$

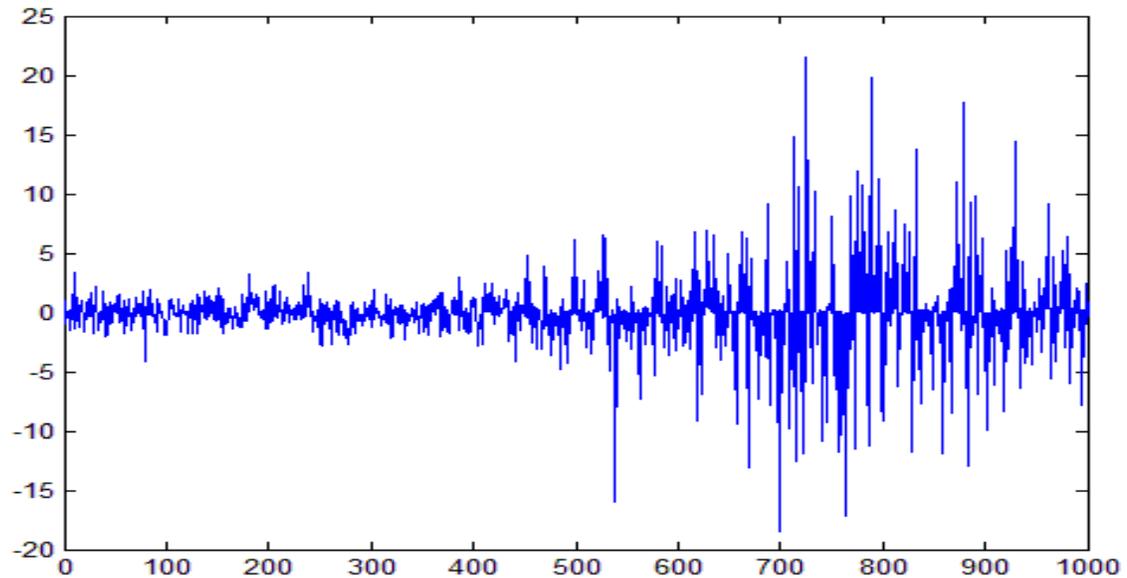


Figure 2.1.3 : Simulation d'une série de taille 1000 du modèle SV de paramètre $\beta = -0.999$
 $\sigma = 0.01$

2.1.3 Etude des propriétés probabilistes du modèle

Stationnarité stricte

Théorème 15 *Le modèle admet, sous la contrainte $|\beta| < 1$, une unique solution strictement stationnaire. Cette solution strictement stationnaire est non anticipative et donnée par*

$$e_t = \exp \left\{ \frac{\omega}{2(1-\beta)} + \frac{\sigma}{2} \sum_{i=0}^{\infty} \beta^i v_{t-i} \right\} \eta_t, \quad t \in \mathbb{Z} \quad (2)$$

Si $|\beta| \geq 1$, il n'existe pas de solution strictement stationnaire non anticipative.

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

Démonstration. Soit $y_t = \log h_t$ d'où l'équation (1) devient :

$$y_t = \omega + \beta y_{t-1} + \sigma v_t \quad (*)$$

En remplaçant $y_{t-1} = \omega + \beta y_{t-2} + \sigma v_{t-1}$ dans (*) on a :

$$y_t = \omega + \beta(\omega + \beta y_{t-2} + \sigma v_{t-1}) + \sigma v_t = [\omega + \beta\omega] + \beta^2 y_{t-2} + [\beta\sigma v_{t-1} + \sigma v_t]$$

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

$$\begin{aligned} y_t &= \omega + \beta\omega + \beta^2(\omega + \beta y_{t-3} + \sigma v_{t-2}) + \beta\sigma v_{t-1} + \sigma v_t \\ &= \omega[1 + \beta + \beta^2] + \beta^3 y_{t-3} + \sigma[\beta^2 v_{t-2} + \beta v_{t-1} + v_t] \end{aligned}$$

Par intuition on peut déduire que :

$$\begin{aligned} y_t &= \omega[1 + \beta + \beta^2 + \beta^3 + \dots] + \sigma[v_t + \beta v_{t-1} + \beta^2 v_{t-2} + \beta^3 y_{t-3} + \dots] \\ &= \frac{\omega}{1 - \beta} + \sigma \sum_{i=0}^{\infty} \beta^i v_{t-i}, \quad \forall t \in \mathbb{Z} \end{aligned}$$

On procède maintenant au raisonnement par récurrence afin de prouver que la relation obtenue est vraie.

Supposons que cette relation (déduite par intuition) soit vraie à un rang de (t-1) c.à.d. :

$$\begin{aligned} y_{t-1} &= \frac{\omega}{1 - \beta} + \sigma \sum_{i=0}^{\infty} \beta^i v_{t-1-i} \\ &= \frac{\omega}{1 - \beta} + \sigma[v_{t-1} + \beta v_{t-2} + \beta^2 v_{t-3} + \beta^3 y_{t-4} + \dots] \end{aligned}$$

Nous allons montrer maintenant qu'elle est vraie au rang t

$$\begin{aligned} y_t &= \frac{\omega}{1 - \beta} + \sigma v_t + \sigma\beta[v_{t-1} + \beta v_{t-2} + \beta^2 v_{t-3} + \dots] \\ &= \sigma v_t + \frac{\omega}{1 - \beta} + \beta[y_{t-1} - \frac{\omega}{1 - \beta}] \\ &= \sigma v_t + \beta y_{t-1} + \omega, \end{aligned}$$

D'où on retrouve la relation (*), en remplaçant $y_t = \log h_t$ (1) on obtient :

$$\left\{ \begin{array}{l} \varepsilon_t = \sqrt{h_t} \eta_t \\ \log h_t = y_t = \frac{\omega}{1-\beta} + \sigma \sum_{i=0}^{\infty} \beta^i v_{t-i} \Rightarrow h_t = \exp\left(\frac{\omega}{1-\beta} + \sigma \sum_{i=0}^{\infty} \beta^i v_{t-i}\right) \\ \text{D'où } \varepsilon_t = \exp\left(\frac{\omega}{2(1-\beta)} + \frac{\sigma}{2} \sum_{i=0}^{\infty} \beta^i v_{t-i}\right) \eta_t \end{array} \right.$$

■

Démonstration. Montrons maintenant la condition de convergence de cette série :

Soit $U_n = \beta^n v_{t-n}$, on dit qu'une série numérique $\sum U_n$ converge absolument lorsque la série des valeurs absolues $\sum |U_n|$ est convergente.

En utilisant le critère de Cauchy c.à.d.

$$\lim_{n \rightarrow \infty} \sqrt[n]{U_n} = l, \text{ telquel } |l| < 1 \Rightarrow \sum U_n \text{ converge}$$

$$\sqrt[n]{U_n} = \sqrt[n]{|\beta^n v_{t-n}|} = \exp\left(\frac{1}{n} \log(|\beta^n v_{t-n}|)\right) = \exp\left(\frac{1}{n} \log |\beta^n| + \frac{1}{n} \log |v_{t-n}|\right) = |\beta| \exp\left(\frac{1}{n} \log |v_{t-n}|\right)$$

$$\lim_{n \rightarrow \infty} \exp\left(\frac{1}{n} \log |v_{t-n}|\right) = \lim_{n \rightarrow \infty} \exp\left(\frac{1}{n} E(\log |v_{t-n}|)\right) = 1$$

car

$$\lim_{n \rightarrow \infty} \exp\left(\frac{1}{n} E(\log |v_{t-n}|)\right) = 0$$

D'où $\lim_{n \rightarrow \infty} \sqrt[n]{U_n} = |\beta|$. Ceci implique que $|\beta| < 1$ pour que la série converge. ■

Stationnarité au second ordre

Théorème 16 Soit $\alpha_i = E\{\exp(\sigma \beta^i v_i)\}$, $i \geq 0$. Si $|\beta| < 1$ et $\prod_{i=0}^{\infty} \alpha_i < \infty$, le processus (ε_t) défini par (2) est un bruit blanc de variance

$$\text{var}(\varepsilon_t) = e^{\frac{\omega}{1-\beta}} \prod_{i=0}^{\infty} \alpha_i$$

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

$$E(\varepsilon_t) = E\left\{\exp\left(\frac{\omega}{2(1-\beta)} + \frac{\sigma}{2} \sum_{i=0}^{\infty} \beta^i v_{t-i}\right)\right\} E(\eta_t) = 0$$

$$\begin{aligned} \text{Var}(\varepsilon_t) &= E\left\{\exp\left(\frac{\omega}{2(1-\beta)} + \frac{\sigma}{2} \sum_{i=0}^{\infty} \beta^i v_{t-i}\right)\right\}^2 E(\eta_t^2) \\ &= \exp\left(\frac{\omega}{1-\beta}\right) \prod_{i=0}^{\infty} E\left\{\exp\left(\frac{\sigma}{2} \beta^i v_{t-i}\right)\right\}^2 \\ &= \exp\left(\frac{\omega}{1-\beta}\right) \prod_{i=0}^{\infty} \alpha_i \end{aligned}$$

De plus, pour tout $k > 0$, $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ en utilisant à nouveau l'indépendance entre (η_t) et (v_t) .

Remarque : Dans le cas où (v_t) est un processus gaussien, il est possible de donner des résultats plus explicites. Si $X \sim N(0, 1)$ on a $E(\exp(\rho X)) = \exp(\frac{\rho^2}{2})$ pour toute constante ρ . Donc $\alpha_i = \exp(\frac{\sigma^2}{2} \beta^{2i})$ et il est clair que la condition $0 < \prod_{i=0}^{\infty} \alpha_i < \infty$ est réalisée si et seulement si $|\beta| < 1$. De plus

$$\text{Var}(\varepsilon_t) = \exp\left(\frac{\omega}{1-\beta} + \frac{\sigma^2}{2(1-\beta^2)}\right)$$

Lorsque (v_t) suit une loi normale centrée réduite.

Le processus (ε_t) ne constitue pas un bruit blanc au sens fort (i.e. un bruit blanc indépendant) ■

Représentation ARMA pour les processus ($\log \varepsilon_t^2$)

En prenant le logarithme du carré de ε_t , la première équation de (1) devient :

$$\log \varepsilon_t^2 = \log h_t + \log \eta_t^2 \quad (3)$$

Cette écriture permet de dériver la fonction d'autocovariance du processus $\log \varepsilon_t^2$. On effectue le changement de notation suivant :

$$X_t = \log \varepsilon_t^2, \quad Y_t = \log h_t \quad \text{et} \quad Z_t = X_t - Y_t = \log \eta_t^2$$

On suppose que Z_t admet des moments jusqu'à l'ordre 4 et on note $\mu_Z = E(Z_t)$, $\sigma_Z^2 = Var(Z_t^2)$. On notera de manière similaire les moments de X_t et Y_t . Le résultat suivant montre que (X_t) admet une représentation ARMA, propriétés qui se révélera très utile pour l'estimation.

Théorème 17 *Si $|\beta| < 1$ et $\sigma_{Z^2}^2 < \infty$, le processus $(X_t) = \log \varepsilon_t^2$ admet une représentation ARMA(1,1) de la forme*

$$X_t - \mu_X = \beta(X_{t-1} - \mu_X) + u_t - \alpha u_{t-1} \quad (4)$$

où (u_t) est un bruit blanc.

Démonstration. Le processus (Y_t) étant solution du modèle AR(1)

$$Y_t = \omega + \beta Y_{t-1} + \sigma v_t$$

ses moyennes et fonction d'autocovariance sont données par

$$\begin{aligned} \mu_Y &= E(Y_t) = \frac{\omega}{1 - \beta} \\ \gamma_Y(0) &= Var(Y_t) = \frac{\sigma^2}{1 - \beta^2} \\ \gamma_Y(k) &= Cov(Y_t, Y_{t-k}) = \beta \gamma_Y(k-1), \quad k > 0 \end{aligned}$$

D'après l'indépendance entre (Y_t) et (Z_t) , (X_t) est un processus du second ordre dont les caractéristiques s'obtiennent comme suit

$$\mu_X = E(X_t) = E(Y_t) + E(Z_t) = \frac{\omega}{1-\beta} + \mu_Z$$

$$\gamma_X(0) = \text{Var}(X_t) = \text{Var}(Y_t) + \text{Var}(Z_t) = \frac{\sigma^2}{1-\beta^2} + \sigma_Z^2$$

$$\gamma_X(k) = \gamma_Y(k) = \beta\gamma_Y(k-1) = \beta^k \frac{\sigma^2}{1-\beta^2}, \quad k > 0$$

Puisque $\gamma_X(k) = \beta\gamma_Y(k-1), \forall k > 0$, (X_t) vérifie une équation ARMA(1,1) de la forme (4). La constante α s'obtient à partir des deux premières autocovariances de (X_t) . D'après (4) on a, en notant σ_u^2 la variance du bruit dans cette représentation

$$\gamma_X(0) = \beta\gamma_X(k-1) + \sigma_u^2\{1 + \alpha(\alpha - \beta)\}$$

$$\gamma_X(1) = -\alpha\sigma_u^2 + \beta\gamma_X(0)$$

Donc si $\sigma_Z^2 \neq 0$

$$\frac{1 + \alpha(\alpha - \beta)}{\alpha} = \frac{\gamma_X(0) - \beta\gamma_X(1)}{\beta\gamma_X(0) - \gamma_X(1)} = \frac{\sigma^2 - \sigma_Z^2}{\beta\sigma_Z^2}$$

Tel que : $\gamma_X(0) = \frac{\sigma^2}{1-\beta^2} + \sigma_Z^2$ et $\gamma_X(1) = \beta\frac{\sigma^2}{1-\beta^2}$ d'où :

$$\alpha = \frac{(1 - \beta^2)\sigma_Z^2 + \sigma^2 - \{(1 + \beta^2)\sigma_Z^2 + \sigma^2\}^{1/2}\{(1 - \beta^2)\sigma_Z^2 + \sigma^2\}^{1/2}}{2\beta\sigma_Z^2}$$

De plus, la variance du bruit dans (4) est donné par

$$\sigma_u^2 = \frac{\beta\sigma_Z^2}{\alpha}$$

Si $\beta \neq 0$ (voir la remarque ci-dessous) (et $\sigma_u^2 = \sigma^2 + \sigma_Z^2$ si $\beta = 0$). Enfin, si $\sigma_Z^2 = 0$ la relation

$$\gamma_X(k) = \beta \gamma_X(k-1)$$

est vraie aussi pour $k = 1$ et (X_t) est un AR(1) (i.e. $\alpha = 0$ dans (4)).

Remarque : Si $\beta = 0$, (X_t) est un bruit blanc fort (indépendant) et $\alpha = 0$. Inversement, si $\beta \neq 0$, on a $\alpha \neq 0$ dès que $\sigma_Z^2 > 0$, et les ordres de la représentation ARMA(1,1) sont exacts. ■

Autocovariance du processus (ε_t^2)

La forme du modèle se prête moins facilement à une étude de la fonction d'autocovariance du processus (ε_t^2) . Celle-ci est utile pour la comparaison avec certaines méthodes d'estimation.

On a en utilisant (2), pour tout $k \geq 0$

$$E(\varepsilon_t^2 \varepsilon_{t-k}^2) = \exp\left(\frac{2\omega}{1-\beta}\right) E\left\{\prod_{i=0}^{\infty} \exp(\sigma \beta^i v_{t-i}) \prod_{i=0}^{\infty} \exp(\sigma \beta^i v_{t-k-i})\right\} E(\eta_t^2 \eta_{t-k}^2)$$

En particulier

$$E(\varepsilon_t^4) = \exp\left(\frac{2\omega}{1-\beta}\right) E\left\{\prod_{i=0}^{\infty} \exp(2\sigma \beta^i v_{t-i})\right\} E(\eta_t^4)$$

Et pour $k > 0$, puisque η_t^2 et η_{t-k}^2 sont indépendantes et d'espérance 1

$$\begin{aligned} E(\varepsilon_t^2 \varepsilon_{t-k}^2) &= \exp\left(\frac{2\omega}{1-\beta}\right) E\left\{\prod_{i=0}^{k-1} \exp(\sigma \beta^i v_{t-i}) \prod_{i=k}^{\infty} \exp(\sigma \beta^i (1 + \beta^{-k}) v_{t-i})\right\} \\ &= \exp\left(\frac{2\omega}{1-\beta}\right) \prod_{i=0}^{k-1} \alpha_i \prod_{i=k}^{\infty} \alpha_{k,i} \end{aligned}$$

En posant, $i \geq 0$, $\alpha_{k,i} = E[\exp(\sigma\beta^i(1 + \beta^{-k})v_t)]$. Donc (ε_t^2) est stationnaire au second ordre dès que

$$|\beta| < 1 \text{ et } E(\eta_t^4) \prod_{i=0}^{\infty} \alpha_{k,i} < \infty$$

Et on a pour tout $k > 0$

$$Cov(\varepsilon_t^2, \varepsilon_{t-k}^2) = \exp\left(\frac{2\omega}{1-\beta}\right) \prod_{i=0}^{k-1} \alpha_i \prod_{i=k}^{\infty} \alpha_{k,i} \prod_{i=0}^{\infty} \alpha_i^2$$

Supposons maintenant que (v_t) suit la loi $N(0, 1)$. On vérifie facilement que $\alpha_{k,i} = \exp(\frac{\sigma^2}{2}\beta^{2i}(1 + \beta^{-k})^2)$ et que la condition de stationnarité au second ordre se limite à $E(\eta_t^4) < \infty$ et $|\beta| < 1$ dans ce cas. Par suite, pour tout $k > 0$

$$\begin{aligned} Cov(\varepsilon_t^2, \varepsilon_{t-k}^2) &= \exp\left(\frac{2\omega}{1-\beta}\right) \left(\prod_{i=0}^{k-1} \exp\left(\frac{\sigma^2}{2}\beta^{2i}\right) \prod_{i=k}^{\infty} \exp\left(\frac{\sigma^2}{2}\beta^{2i}(1 + \beta^{-k})^2\right) - \prod_{i=0}^{\infty} \exp(\sigma^2\beta^{2i}) \right) \\ &= \exp\left(\frac{2\omega}{1-\beta} + \frac{\sigma^2}{1-\beta^2}\right) \left(\exp\left(\frac{\sigma^2}{1-\beta^2}\beta^k - 1\right) \right) \end{aligned}$$

Il est intéressant de remarquer que la fonction d'autocovariance de (ε_t^2) tend vers zéro lorsque k tend vers l'infini mais que la décroissance n'est pas compatible avec l'équation de récurrence linéaire reliant les autocovariances. Or une telle relation est caractéristique des modèles ARMA. On peut donc en déduire que (ε_t^2) , bien que stationnaire au second ordre, n'admet pas de représentation ARMA. On se retrouve ici dans un cas où la représentation de Wold ne se réduit pas à une écriture ARMA. Cependant on a un équivalent de forme $Cov(\varepsilon_t^2, \varepsilon_{t-k}^2) \sim cste \times \beta^k$ lorsque $k \rightarrow \infty$, ce qui montre que la vitesse asymptotique de décroissance (exponentielle) des autocovariances est la même que pour un processus ARMA.

Chapitre 3

Estimation du modèle de volatilité stochastique

3.1 Méthode du Pseudo-Maximum de Vraisemblance

3.1.1 Introduction

Le caractère latent (inobservable) de la volatilité rend l'inférence statistique des modèles à volatilité stochastique plus difficile. En particulier la loi conditionnelle du processus observé n'est pas explicite, ce qui empêche d'écrire la fonction de vraisemblance sous une forme simple permettant de la maximiser.

$$\begin{cases} \varepsilon_t = \sqrt{h_t} \eta_t \\ \log h_t = \omega + \beta \log h_{t-1} + \sigma v_t \end{cases} \quad (3.1)$$

Notons que $\theta = (\omega, \beta, \alpha)$ le vecteur des paramètres et θ_0 la vraie valeur.

La taille de l'échantillon est n , soit $\varepsilon_1, \dots, \varepsilon_n$, du processus (ε_t) . Rappelons que les autres variables du modèle, en particulier h_t , ne sont pas observables.

3.1.2 Présentation de la méthode

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

blance s'écrit donc :

$$l(\varepsilon_1, \dots, \varepsilon_n; \theta) = \prod_{t=2}^n f_t(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1}) f_1(\varepsilon_1)$$

Où, pour $2 \leq t \leq n$, $f_t(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1})$ désigne la densité de ε_t conditionnelle à $(\varepsilon_1, \dots, \varepsilon_{t-1})$ et $f_1(\cdot)$ la densité de ε_1 .

Lorsque les lois conditionnelles sont gaussiennes cette décomposition permet d'obtenir une écriture simple de la log-vraisemblance en fonction des moyennes et variance conditionnelles des variables ε_t .

En considérant $\varepsilon_1 \sim N(0, \sigma_1^2(\theta))$ suit un processus gaussien, on peut donc écrire la densité de la première observation ε_1 comme suit :

$$f_1(\varepsilon_1) = \frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}} \exp\left(-\frac{\varepsilon_1^2}{2\sigma_1^2(\theta)}\right) \quad (3.2)$$

Supposons maintenant que le processus (ε_t) conditionné par rapport à son passé soit gaussien, c.à.d. $(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1}) \sim N(m_t(\theta), \sigma_t^2(\theta))$, telle que $m_t(\theta) = E(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1})$ et $\sigma_t^2(\theta) = Var(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1})$.

Sachant que le processus $(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1})$ est gaussien implique que sa distribution s'écrit :

$$f(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left(-\frac{(\varepsilon_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right) \quad (3.3)$$

On peut donc écrire la log-vraisemblance

$$\begin{aligned}
 \log(l(\varepsilon_1, \dots, \varepsilon_n; \theta)) &= \log(f_1(\varepsilon_1)) + \sum_{t=2}^n \log f_t(\varepsilon_t \mid \varepsilon_1, \dots, \varepsilon_{t-1}) \\
 &= \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}} \exp\left(-\frac{\varepsilon_1^2}{2\sigma_1^2(\theta)}\right)\right) + \sum_{t=2}^n \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left(-\frac{(\varepsilon_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right)\right) \\
 &= \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2(\theta)}}\right) + \sum_{t=2}^n \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}}\right) + \left(-\frac{\varepsilon_1^2}{2\sigma_1^2(\theta)}\right) + \sum_{t=2}^n \left(-\frac{(\varepsilon_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right) \\
 &= \sum_{t=1}^n \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}}\right) + \sum_{t=1}^n \left(-\frac{(\varepsilon_t - m_t(\theta))^2}{2\sigma_t^2(\theta)}\right) \\
 &= -\sum_{t=1}^n \log \sqrt{2\pi} - \sum_{t=1}^n \log(\sqrt{\sigma_t^2(\theta)}) - \frac{1}{2} \sum_{t=1}^n \left(\frac{(\varepsilon_t - m_t(\theta))^2}{\sigma_t^2(\theta)}\right) \\
 &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log(\sigma_t^2(\theta)) + \frac{(\varepsilon_t - m_t(\theta))^2}{\sigma_t^2(\theta)}\right)
 \end{aligned}$$

Modèle Espace Etat

La méthode de vraisemblance consiste à trouver une approximation de la vraisemblance, obtenue à partir du filtre de Kalman. Celui-ci s'applique aux modèles admettant une représentation *espace-état* et peut être utilisé pour le filtrage, le lissage et la prédiction.

L'écriture générale d'une représentation espace-état se présente sous la forme de deux équations :

$$\begin{cases} y_t = U_t \alpha_t + d_t + u_t \\ \alpha_t = T_t \alpha_{t-1} + c_t + R_t v_t \end{cases} \quad t \geq 1$$

La première équation est appelée *équation de mesure* et la seconde *équation de transition*. La variable y_t est observable et à valeurs dans \mathbb{R}^N tandis que les m composantes du vecteur α_t , appelé *vecteur d'état*, sont généralement supposées non observables. Les matrices U_t , d_t , T_t , c_t et R_t sont supposées non stochastiques mais peuvent dépendre de t . Les suites (u_t) et

(v_t) sont centrées, non corrélées (c.à.d. $E(u_t v_s) = 0 \forall t, s$) mais leur variance peut dépendre du temps.

La forme d'espace d'état du modèle (1) est obtenue par passage au logarithme :

$$\begin{cases} \varepsilon_t^2 = h_t \eta_t^2 \\ \log h_t = \omega + \beta \log h_{t-1} + \sigma v_t \end{cases} \Leftrightarrow \begin{cases} \log(\varepsilon_t^2) = \log(h_t) + \log(\eta_t^2) \\ \log h_t = \beta \log h_{t-1} + \omega + \sigma v_t \end{cases}$$

En posant, $Z_t = \log(\eta_t^2)$, $m_Z = E(Z_t)$ et $u_t = Z_t - m_Z$

$$\begin{cases} \log(\varepsilon_t^2) = \log(h_t) + Z_t + m_Z - m_Z \\ \log h_t = \beta \log h_{t-1} + \omega + \sigma v_t \end{cases} \Leftrightarrow \begin{cases} \log(\varepsilon_t^2) = \log(h_t) + m_Z + u_t \\ \log h_t = \beta \log h_{t-1} + \omega + \sigma v_t \end{cases}$$

On note également que $\sigma_Z^2 = Var(\log(\eta_t^2))$, en supposant que cette quantité existe.

Application du Filtre de Kalman

Soit

$$\alpha_{t|t} = E(\alpha_t | y_1, \dots, y_t), \quad P_{t|t} = Var(\alpha_t | y_1, \dots, y_t)$$

$$\alpha_{t|t-1} = E(\alpha_t | y_1, \dots, y_{t-1}),$$

$$P_{t|t-1} = Var(\alpha_t | y_1, \dots, y_{t-1}) = E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})']$$

Comme il s'agit ici de calculer l'estimation du vecteur d'état sur la base des données observées, donc $\alpha_{t|t-1}$ représente la projection de α_t sur (y_1, \dots, y_{t-1}) . Sachant que le filtre de Kalman calcule récursivement successivement les estimateurs $\alpha_{1|0}, \alpha_{2|1}, \dots, \alpha_{t|t-1}$, alors on associe à chaque estimateur une matrice de covariance de l'erreur de l'estimation de α_t , représentée par $P_{t|t-1}$.

Première étape : en prenant l'espérance conditionnelle par rapport à (y_1, \dots, y_{t-1}) dans l'équation de transition, on obtient :

$$\begin{aligned}
 \alpha_{t|t-1} &= E(\alpha_t \mid y_1, \dots, y_{t-1}) = E((T_t \alpha_{t-1} + c_t) \mid (y_1, \dots, y_{t-1})) \\
 &= T_t E(\alpha_{t-1} \mid y_1, \dots, y_{t-1}) + c_t \\
 \alpha_{t|t-1} &= T_t \alpha_{t-1|t-1} + c_t
 \end{aligned}$$

puis, en utilisant la relation $\alpha_t - \alpha_{t|t-1} = T_t(\alpha_{t-1} - \alpha_{t-1|t-1}) + R_t v_t$, afin de simplifier le calcul de la matrice $P_{t|t-1}$.

$$\begin{aligned}
 P_{t|t-1} &= E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})'] \\
 &= E[(T_t(\alpha_{t-1} - \alpha_{t-1|t-1}) + R_t v_t)(T_t(\alpha_{t-1} - \alpha_{t-1|t-1}) + R_t v_t)'] \\
 &= T_t E[(\alpha_{t-1} - \alpha_{t-1|t-1})(\alpha_{t-1} - \alpha_{t-1|t-1})'] T_t' + R_t E(v_t v_t') R_t' \\
 &= T_t P_{t-1|t-1} T_t' + R_t Q_t R_t'
 \end{aligned}$$

Les autres termes s'annulent à cause de l'indépendance entre α_t et v_t condition de normalité (4). Ces équations sont appelées équations de prévision.

Deuxième étape : une fois disponible l'observation y_t , les quantités précédentes sont mises à jour :

$$\alpha_{t|t} = E(\alpha_t \mid y_1, \dots, y_t)$$

Ceci peut être évalué en utilisant la formule de mise à jour de la projection linéaire de α_t sur l'ensemble des observations (y_1, \dots, y_t) c.à.d.

$$\alpha_{t|t} = \alpha_{t|t-1} + E[(\alpha_t - \alpha_{t|t-1})(y_t - y_{t|t-1})'] E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})']^{-1} (y_t - y_{t|t-1})$$

Avant de calculer $\alpha_{t|t}$ on doit d'abord calculer l'innovation de y_t à l'instant t ($y_t - y_{t|t-1}$) :

telle que : $y_{t|t} = E(y_t \mid \alpha_t, v_t) = E(U_t \alpha_t + d_t + u_t \mid \alpha_t, v_t) = U_t \alpha_{t|t} + d_t$ d'où

$$y_t - y_{t|t-1} = U_t(\alpha_t - \alpha_{t|t-1}) + u_t$$

En lui associant sa matrice de covariance :

$$\begin{aligned}
 E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})'] &= E[(U_t(\alpha_t - \alpha_{t|t-1}) + u_t)(U_t(\alpha_t - \alpha_{t|t-1}) + u_t)'] \\
 &= U_t E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})'] U_t' + E(u_t u_t') \\
 &= U_t P_{t|t-1} U_t' + H_t
 \end{aligned}$$

En posant, F_t : matrice de gain de kalman

$$F_t = Var(y_t | y_1, \dots, y_{t-1}) = E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})'] = U_t P_{t|t-1} U_t' + H_t$$

Une fois la matrice de covariance obtenue, on peut maintenant calculer $\alpha_{t|t}$:

$$\begin{aligned}
 \alpha_{t|t} &= \alpha_{t|t-1} + E[(\alpha_t - \alpha_{t|t-1})(y_t - y_{t|t-1})'] E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})']^{-1} (y_t - y_{t|t-1}) \\
 &= \alpha_{t|t-1} + E[(\alpha_t - \alpha_{t|t-1})(U_t(\alpha_t - \alpha_{t|t-1}) + u_t)'] F_t^{-1} (y_t - y_{t|t-1}) \\
 &= \alpha_{t|t-1} + E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})'] U_t' F_t^{-1} (y_t - y_{t|t-1}) \\
 &= \alpha_{t|t-1} + P_{t|t-1} U_t' F_t^{-1} (y_t - U_t \alpha_{t|t-1} - d_t)
 \end{aligned}$$

Enfin, on retrouve la matrice de covariance de l'erreur de prédiction, notée $P_{t|t}$, comme suit :

$$\begin{aligned}
 P_{t|t} &= E[(\alpha_t - \alpha_{t|t})(\alpha_t - \alpha_{t|t})'] \\
 &= E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})'] - \{E[(\alpha_t - \alpha_{t|t-1})(y_t - y_{t|t-1})']\} \{E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})']\}^{-1} \\
 &\quad \{E[(y_t - y_{t|t-1})(\alpha_t - \alpha_{t|t-1})']\} \\
 &= P_{t|t-1} - P_{t|t-1} U_t' F_t^{-1} U_t P_{t|t-1}
 \end{aligned}$$

On applique maintenant l'algorithme :

Initialisation de la première étape : $\alpha_{1|0} = T_{1|0} + c_1$, $P_{1|0} = T_1 P_0 T_1' + R_1 Q_1 R_1'$

On calcule ainsi d'une manière récursive pour $t = 1, \dots, n$, les quantités $\alpha_{t|t-1}$, $P_{t|t-1}$ et $\alpha_{t|t}$, $P_{t|t}$. Ceci permet en même temps d'obtenir $E(y_t | y_1, \dots, y_{t-1}) = U_t \alpha_{t|t-1} + d_t$, $F_t = Var(y_t | y_1, \dots, y_{t-1})$ et donc d'écrire la log-vraisemblance gaussienne du modèle.

$$\text{Soit } \alpha_{t|t} = E(\log h_t | \varepsilon_1^2, \dots, \varepsilon_t^2), P_{t|t} = Var(\log h_t | \varepsilon_1^2, \dots, \varepsilon_t^2)$$

$$\alpha_{t|t-1} = E(\log h_t | \varepsilon_1^2, \dots, \varepsilon_{t-1}^2), P_{t|t-1} = Var(\log h_t | \varepsilon_1^2, \dots, \varepsilon_{t-1}^2)$$

L'algorithme prend ici la forme simplifiée, en remplaçant ($y_t = \log(\varepsilon_t^2)$), $U_t = 1$, $\alpha_t = \log h_t$, $d_t = m_Z$, $T_t = \beta$, $c_t = \omega$, $R_T = \sigma$, $Q_t = 1$) dans les équations précédentes on retrouve :

$$\alpha_{t|t-1} = T_t \alpha_{t-1|t-1} + c_t = \beta \alpha_{t-1|t-1} + \omega$$

$$P_{t|t-1} = T_t P_{t-1|t-1} T_t' + R_t Q_t R_t = \beta^2 P_{t-1|t-1} + \sigma^2$$

$$\alpha_{t|t} = \alpha_{t|t-1} + P_{t|t-1} U_t' F_t^{-1} (y_t - U_t \alpha_{t|t-1} - d_t) = \alpha_{t|t-1} + P_{t|t-1} F_t^{-1} (\log(\varepsilon_t^2) - \alpha_{t|t-1} - m_Z)$$

$$\alpha_{1|0} = \frac{\omega}{1 - \beta}, \quad P_{1|0} = \frac{\sigma^2}{(1 - \beta^2)}$$

et

$$F_t = Var(\log(\varepsilon_t^2) | \log(\varepsilon_1^2), \dots, \log(\varepsilon_{t-1}^2)) = U_t P_{t|t-1} U_t' + H_t = P_{t|t-1} + \sigma_Z^2$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} U_t' F_t^{-1} U_t P_{t|t-1} = P_{t|t-1} - P_{t|t-1}^2 F_t^{-1}$$

D'où finalement la (pseudo) log-vraisemblance :

$$\log(l(\varepsilon_1, \dots, \varepsilon_n; \theta)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log(\sigma_t^2(\theta)) + \frac{(\varepsilon_t - m_t(\theta))^2}{\sigma_t^2(\theta)})$$

$$\log(l(\varepsilon_1, \dots, \varepsilon_n; \theta)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log(F_t) + \frac{(\log(\varepsilon_t^2) - \alpha_{t|t-1} - m_Z)^2}{F_t})$$

Remarque : Pour des valeurs fixées de $m_Z = -1.27$ et $\sigma_Z^2 = \pi^2/2$ (d'après Abramovitz et stegun 1970), le vecteur θ des paramètres (ω, β, σ) peut alors être estimé en maximisant cette fonction, construite en appliquant l'algorithme précédent.

Procédure de Lissage

Une fois, le filtre de Kalman est calculé et les valeurs des suites $\{\alpha_{t|t}\}_{t=1}^T$ $\{\alpha_{t|t-1}\}_{t=2}^T$ $\{P_{t|t}\}_{t=1}^T$ $\{P_{t|t-1}\}_{t=2}^T$ sont stockées.

c.à.d.

$$\alpha_{t|t-1} = T_t \alpha_{t-1|t-1} + c_t = \beta \alpha_{t-1|t-1} + \omega$$

$$P_{t|t-1} = T_t P_{t-1|t-1} T_t' + R_t Q_t R_t = \beta^2 P_{t-1|t-1} + \sigma^2$$

L'estimation lissée de la dernière valeur de l'échantillon $\alpha_{T|T}$ est la dernière entrée dans $\{\alpha_{t|t}\}_{t=1}^T$. Alors, on utilise la matrice de gain $\{J_t\}_{t=1}^T$

$$J_{t-1} = P_{t-1|t-1} T_t' P_{t|t-1}^{-1}$$

comme générateur, et à partir des équations précédentes et en remplaçant

($y_t = \log(\varepsilon_t^2)$), $U_t = 1$, $\alpha_t = \log h_t$, $d_t = m_Z$, $T_t = \beta$, $c_t = \omega$, $R_T = \sigma$, $Q_t = 1$) on retrouve :

$$J_{t-1} = P_{t-1|t-1} T_t' P_{t|t-1}^{-1} = P_{t-1|t-1} \beta P_{t|t-1}^{-1}$$

$$\alpha_{t|T} = \alpha_{t|t} + J_t (\alpha_{t+1|T} - \alpha_{t+1|t})$$

$$P_{t|T} = P_{t|t} + J_t (P_{t+1|T} - P_{t+1|t}) J_t'$$

3.1.3 Estimation des paramètres

Afin d'estimer les paramètres du modèle SV, nous avons testé ce modèle sur des séries d'observations (Y_n) de taille T variant de 500 à 6000 générées à partir de l'équation du modèle canonique. Ces séries sont générées en attribuant dans chaque essai des valeurs initiales des paramètres $(\omega, \beta, \sigma^2)$. Pour estimer les paramètres du modèle, nous avons utilisé l'algorithme du pseudo maximum de vraisemblance, implémenté sous l'environnement Matlab 7.0. Le tableau suivant représente l'estimation de chaque paramètre ainsi que son écarts-type entre parenthèse.

Tableau 1 : Expériences Monte Carlo sur les estimateurs de la méthode QML, d'un modèle stationnaire de volatilité stochastique : Rprésentant la valeur des estimateurs avec leurs écarts types (entre parenthèses

	T = 500	T = 3000	T = 6000
$\beta = 0.9$	0.8982 (0.0299)	0.8817 (0.0132)	0.8997 (0.0081)
$\sigma^2 = 1$	0.7336 (0.1013)	0.9867 (0.0503)	0.9921 (0.0346)
$\beta = 0.7$	0.6785 (0.1310)	0.6792 (0.0390)	0.6960 (0.0272)
$\sigma^2 = 1$	0.8722 (0.2233)	1.0681 (0.0785)	0.9831 (0.0541)
$\beta = 0.9$	0.8052 (0.1461)	0.9186 (0.0256)	0.8994 (0.0198)
$\sigma^2 = 0.09$	0.3220 (0.1819)	0.2809 (0.0573)	0.2925 (0.0386)
$\beta = 0.7$	0.6280 (0.3156)	0.8025 (0.1695)	0.7113 (0.1369)
$\sigma^2 = 0.09$	0.5512 (0.3183)	0.2599 (0.1603)	0.2691 (0.0993)

Les résultats du tableau montrent bien que les valeurs estimées sont proches des valeurs réelles et cela en augmentant à chaque fois la taille de la série ce qui confirme la convergence de la méthode du maximum de vraisemblance vers la valeur réelle des paramètres.

L'exemple suivant représente une série de paramètres ($\beta = 0.9, \sigma^2 = 1$) de taille $T = 500$

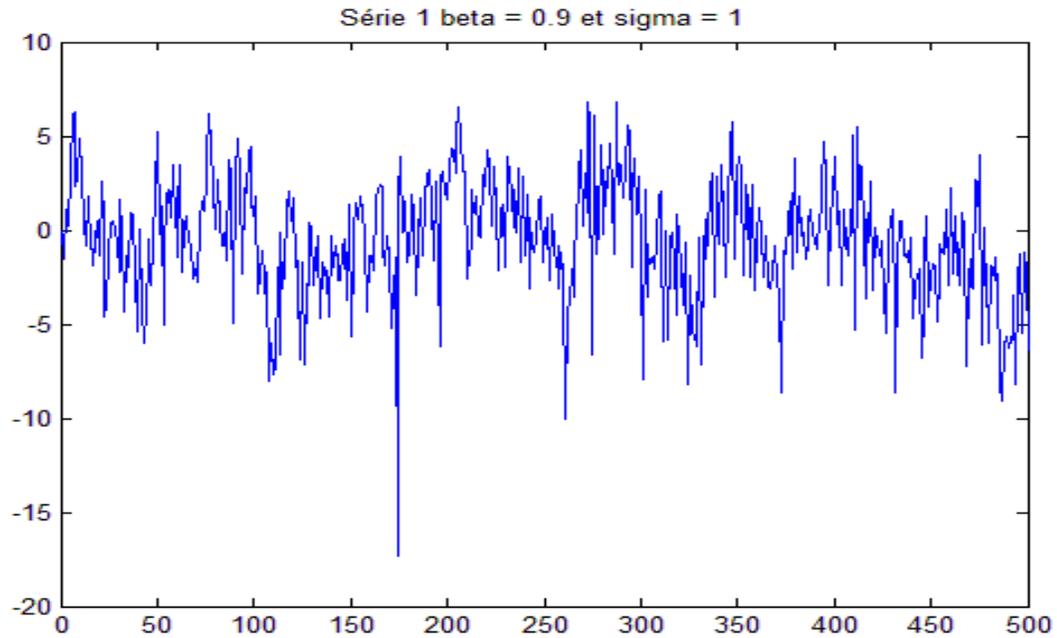


Figure 3.1.1 : Série générée par ($\beta = 0.9, \sigma^2 = 1$) et $T = 500$.

La valeur des paramètres après estimation est ($\hat{\beta} = 0.8829(0.0326), \hat{\sigma}^2 = 1.0184(0.1271)$), d'où nous obtenons le graphe de la log-volatilité ($\log(h_t)$) .

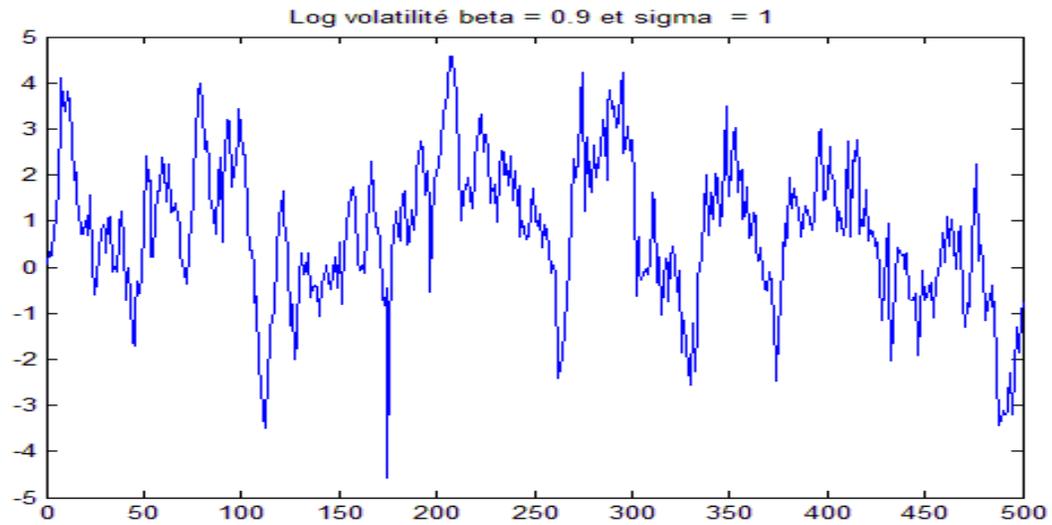


Figure 3.1.2 : Log-volatilité de la série ($\beta = 0.9$ $\sigma^2 = 1$) et $T = 500$.

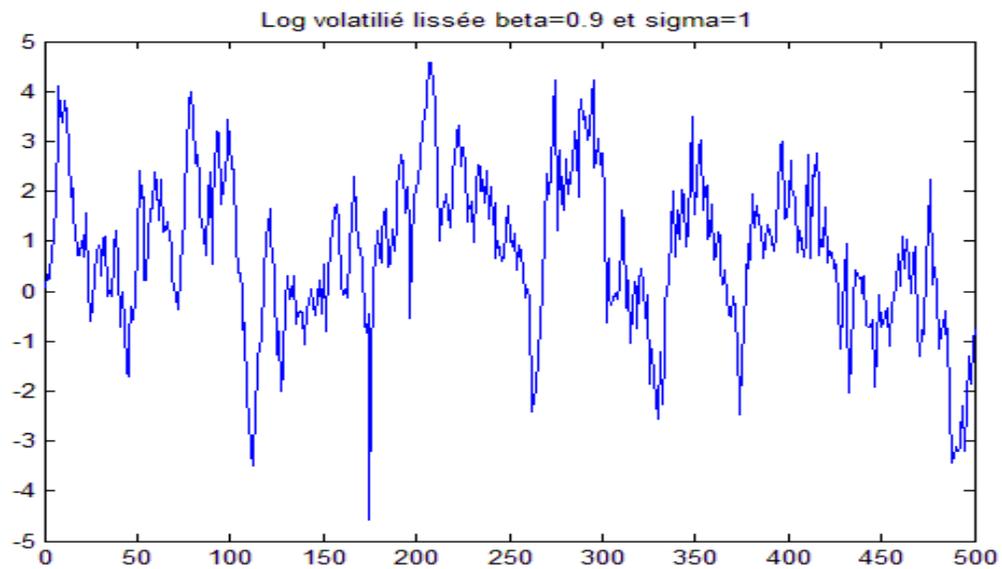


Figure 3.1.3 : Lissage de la log-volatilité

3.2 Méthode de Simulation Monte Carlo par Chaînes de Markov (MCMC)

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.

3.2.2 Présentation de la méthode

L'objectif de l'analyse bayésienne revient à calculer l'expression suivante :

$$E(h(\theta)/X) = \int_{\theta} h(\theta)p(\theta/X)d\theta \quad (1)$$

Avec X est le vecteur représentant les données , θ vecteur des paramètres et $h(\theta)$ une fonction quelconque où l'espérance existe.

Le problème qui se pose en général, c'est qu'il est difficile de trouver une solution analytique à cette équation. Pour résoudre ce problème, on a recours à des algorithmes de simulations qui relèvent de la Chaîne Markovienne de Monte Carlo, en particulier l'algorithme de Metropolis-Hasting. Ce genre d'algorithme est utilisé généralement dans le cas où la densité a priori et a posteriori ne sont pas conjuguées.

La méthode générale de Monte Carlo consiste à tirer un $\theta^{(m)}$ pour $m = 1...M$ de $p(\theta/X)$ et utiliser le fait que lorsque $M \rightarrow \infty$

$$\frac{1}{M} \sum_{m=1}^M h(\theta^{(m)}) \xrightarrow{ps} E(h(\theta)/X) \quad (2)$$

La méthode de MCMC consiste à construire un processus markovien $p(\theta^{(m+1)}/\theta^{(m)})$ qui assurent la convergence vers la loi stationnaire unique $p(\theta/X)$.

L'algorithme de Metropolis-Hasting qui relève de la MCMC sert à construire une chaîne de Markov dont la loi stationnaire et unique est la loi a posteriori.

Cet algorithme peut être présenté brièvement comme suit :

- Choisir une valeur initiale $\theta^{(0)}$.
- Tirer un candidat $\theta^{(*)}$ selon une loi de proposition.
- Calculer le ratio de Hasting $\alpha(\theta^{(m-1)}, \theta^*)$, définit par la probabilité :

$$\alpha(\theta^{(m-1)}, \theta^*) = \min \left(\frac{p(\theta = \theta^*/X)}{p(\theta = \theta^{(m-1)}/X)}, 1 \right)$$

- Tirer une variable aléatoire u d'une loi uniforme $U[0, 1]$ et la comparer au ratio de Hasting $\alpha(\theta^{(m-1)}, \theta^*)$.

Si $u < \alpha$, $\theta^{(m)} = \theta^*$, le candidat est accepté

Si non $\theta^{(m)} = \theta^{(m-1)}$, le candidat est rejeté

- Répéter les trois étapes précédentes M fois.
- Utiliser l'équation (2) afin d'approximer $E(h(\theta)/X)$ pour n'importe quelle fonction $h(\theta)$ où l'espérance existe.

La règle d'acceptation ou du rejet du candidat sert généralement à s'assurer que la chaîne se déplace dans la bonne direction.

Cet algorithme va nous servir de pour simuler la dynamique des volatilités latentes.

3.2.3 Estimation des paramètres

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

$$\begin{cases} y_t = \exp(h_t/2)\epsilon_t \\ h_t = \mu + \phi h_{t-1} + \tau\eta_t \end{cases} \quad (1)$$

où ϵ_t et η_t suivent $N(0, 1)$.

Soit $y^n = (y_1, \dots, y_n)$, $h^n = (h_1, \dots, h_n)$ et $h_{a:b} = (h_a, \dots, h_b)$.

Densité a priori

La densité a priori $f(\theta)$ est la fonction de densité des paramètres avant d'intégrer les données.

L'incertitude concernant la log-volatilité initiale est $h_0 \sim N(m_0, C_0)$

Soit $\theta = (\mu, \phi)$ vecteur des paramètres, alors la distribution a priori de (θ, τ^2) est une normale-inverse gamma, i.e.

$$(\theta, \tau^2) \sim NIG(\theta_0, V_0, \nu_0, s_0^2)$$

$$\theta/\tau^2 \sim N(\theta_0, \tau^2 V_0)$$

$$\tau^2 \sim IG(\nu_0/2, \nu_0 s_0^2/2)$$

Les hyperparamètres sont : $m_0, C_0, \theta_0, V_0, \nu_0$ et s_0^2 .

Lois a posteriori

La distribution a posteriori de notre modèle via la méthode MCMC peut être effectuée par au moins deux façons.

Soit $h_{-t} = (h_{0:(t-1)}, h_{(t+1):n})$, pour $t = 1, \dots, n-1$ et $h_{-n} = h_{1:(n-1)}$

Choix d'échantillon $(\theta, \tau^2/h^n, y^n)$

La distribution a posteriori de (θ, τ^2) conditionnée sur $h_{0:n}$ suit aussi une normale-inverse gamma :

$$(\theta, \tau^2)/y^n, h_{0:n} \sim NIG(\theta_1, V_1, \nu_1, s_1^2)$$

où $X = (1_n, h_{0:(n-1)})$, $\nu_1 = \nu_0 + n$

$$V^{-1} = V_0^{-1} + X'X$$

$$V_1^{-1}\theta_1 = V_0^{-1}\theta_0 + X'h_{1:n}$$

$$\nu_1 s_1^2 = \nu_0 s_0^2 + (y - X\theta_1)'(y - X\theta_1) + (\theta_1 - \theta_0)'V_0^{-1}(\theta_1 - \theta_0)$$

Choix d'échantillon $(h_0/\theta, \tau^2, h_1)$

En combinant

$$h_0 \sim N(m_0, C_0)$$

et

$$h_1/h_0 \sim N(\mu + \phi h_0, \tau^2)$$

permet d'obtenir (par le théorème de Bayes)

$$h_0/h_1 \sim N(m_1, C_1)$$

où

$$C_1^{-1}m_1 = C_0^{-1}m_0 + \phi\tau^{-2}(h_1 - \mu)$$

$$C_1^{-1} = C_0^{-1} + \phi^2\tau^{-2}$$

Distribution a priori conditionnée sur h_t

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

$$\begin{pmatrix} h_t \\ h_{t+1} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu + \phi h_{t-1} \\ (1 + \phi)\mu + \phi^2 h_{t-1} \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & \phi \\ \phi & (1 + \phi^2) \end{pmatrix} \right\}$$

où $E(h_t/h_{t-1}, h_{t+1}, \theta, \tau^2)$ est

$$\mu_t = \left(\frac{1 - \phi}{1 + \phi^2} \right) \mu + \left(\frac{\phi}{1 + \phi^2} \right) (h_{t-1} + h_{t+1})$$

et $V(h_t/h_{t-1}, h_{t+1}, \theta, \tau^2)$ est

$$\nu^2 = \tau^2(1 + \phi^2)^{-1}$$

Par conséquent,

$$(h_t/h_{t-1}, h_{t+1}, \theta, \tau^2) \sim N(\mu_t, \nu^2) \quad t = 1, \dots, n - 1$$

$$(h_n/h_{n-1}, \theta, \tau^2) \sim N(\mu_n, \tau^2)$$

$$\text{où } \mu_n = \mu + \phi h_{n-1}$$

Etablir un échantillon h_t via l'algorithme Metropolis-Hastings

La distribution conditionnelle de h_t est donnée par :

$$\begin{aligned} p(h_t/h_{-t}, y^n, \theta, \tau^2) &= p(h_t/h_{t-1}, h_{t+1}, \theta, \tau^2) p(y_t/h_t) \\ &= f_N(h_t; \mu_t, \nu^2) f_N(y_t; 0, e^{h_t}) \end{aligned}$$

Algorithme de Metropolis-Hastings

Pour $t = 1, \dots, n$

1. L'état actuel : $h_t^{(j)}$
2. Etablir un échantillon h_t^* à partir d'une loi normale $N(\tilde{\mu}_t, \nu_t^2)$
3. Calculer la probabilité d'acceptation

$$\alpha = \min \left\{ 1, \frac{f_N(h_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^*})}{f_N(h_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^{(j)}})} \times \frac{f_N(h_t^{(j)}; \tilde{\mu}_t, \nu_t^2)}{f_N(h_t^*; \tilde{\mu}_t, \nu_t^2)} \right\}$$

4. Nouvel état :

$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{avec la probabilité } \alpha \\ h_t^{(j)} & \text{avec la probabilité } 1 - \alpha \end{cases}$$

Soit l'exemple suivant :

On veut estimer les paramètres du modèle 1 en simulant une série de taille $n = 500$, nombre d'itération Monte Carlo est $N = 4000$ avec les paramètres suivants $\mu = -0.0065$, $\phi = 0.99$ et $\tau^2 = 0.15^2$

Les graphes suivants représentent la simulation de la série ainsi que sa volatilité

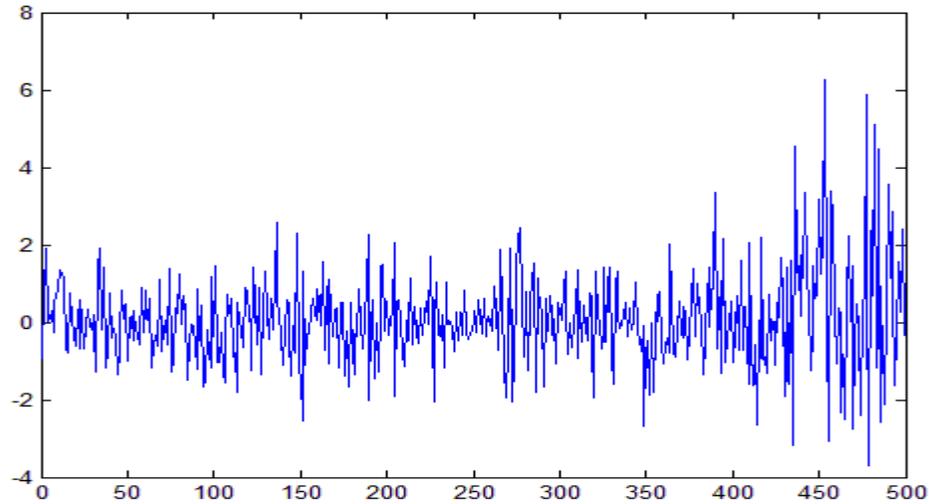


Figure 3.2.1 : Estimation Série taille $n = 500$, et $\mu = -0.0065$ $\phi = 0.99$ et $\tau^2 = 0.15^2$

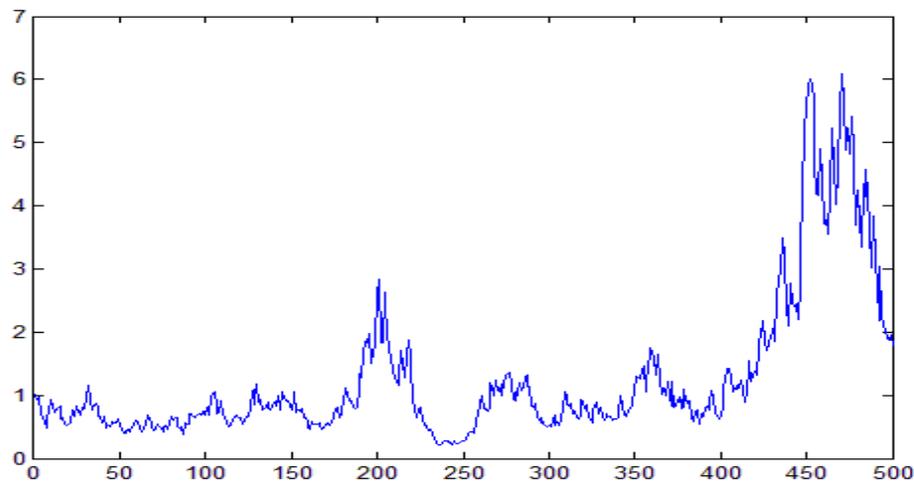


Figure 3.2.2 : Volatilté de la Série taille $n = 500$, et $\mu = -0.0065$ $\phi = 0.99$ et $\tau^2 = 0.15^2 = 0.0225$

Le tableau suivant résume les valeurs des paramètres estimés est :

Paramètres	μ	ϕ	τ^2
Valeurs initiales	-0.0065	0.99	0.0225
Estimation	0.0019	0.9737	0.0299

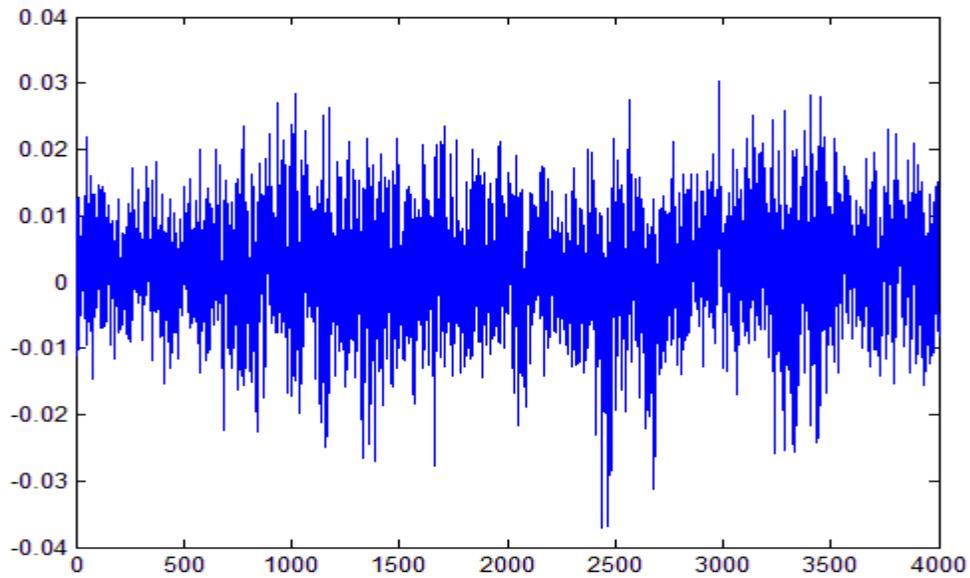


Figure 3.2.3 : Valeurs estimées de μ sur 4000 itérations

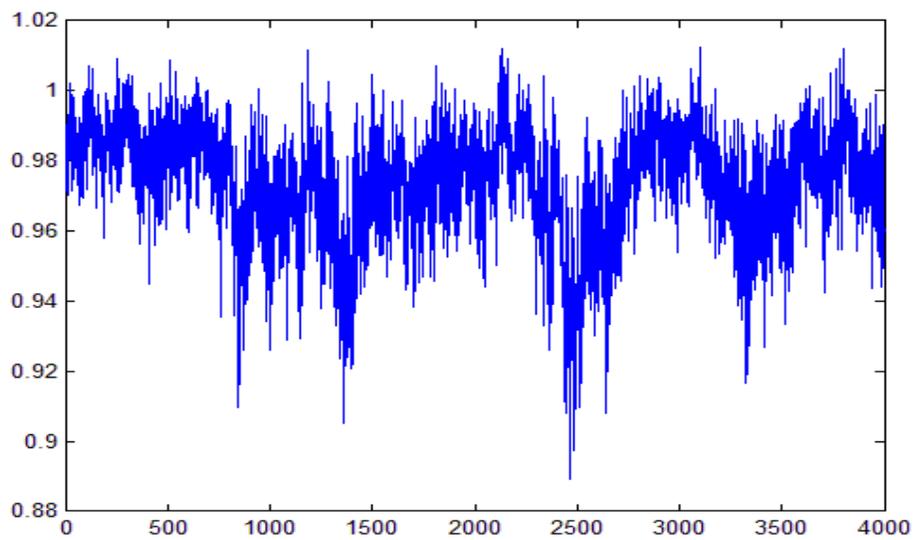


Figure 3.2.4 : Valeurs estimées de ϕ sur 4000 itérations

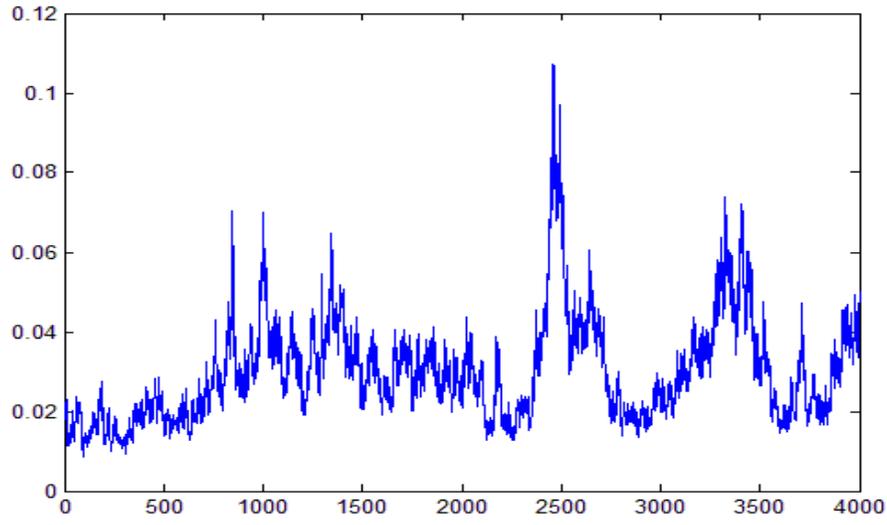


Figure 3.2.5 : Valeurs estimées de τ^2 sur 4000 itérations

Le graphe suivant représente la volatilité correspondant à la série de la figure 1.

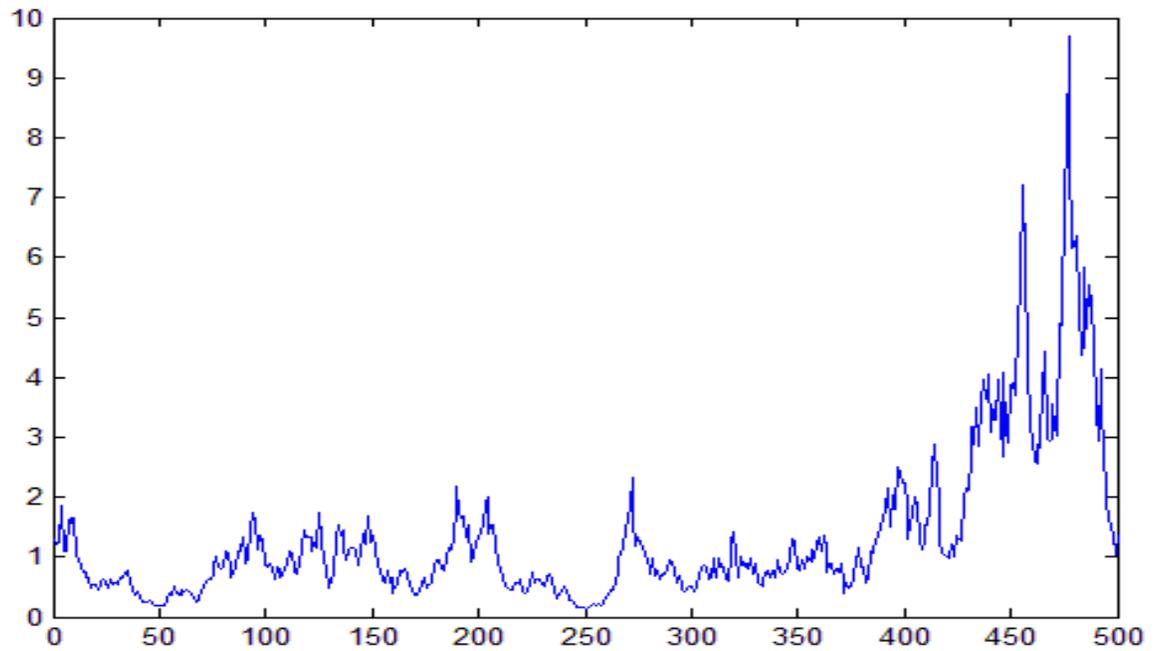


Figure 3.2.6 : Volatilité avec les paramètres $\hat{\mu} = 0.0019$ $\hat{\phi} = 0.9737$ et $\hat{\tau}^2 = 0.0299$

3.3 Application du modèle SV sur un cas réel

3.3.1 Introduction

Pour tester la pertinence des deux méthodes utilisées à savoir, la méthode QML et la méthode MCMC, nous avons appliqué au modèle SV, une série financière montrant les rendements quotidiens de la Bourse de New York (New York Stock Exchange (NYSE)) à partir du 2 Février 1984 jusqu'au 31 Décembre 1991. Cette série est représentée dans la figure suivante.

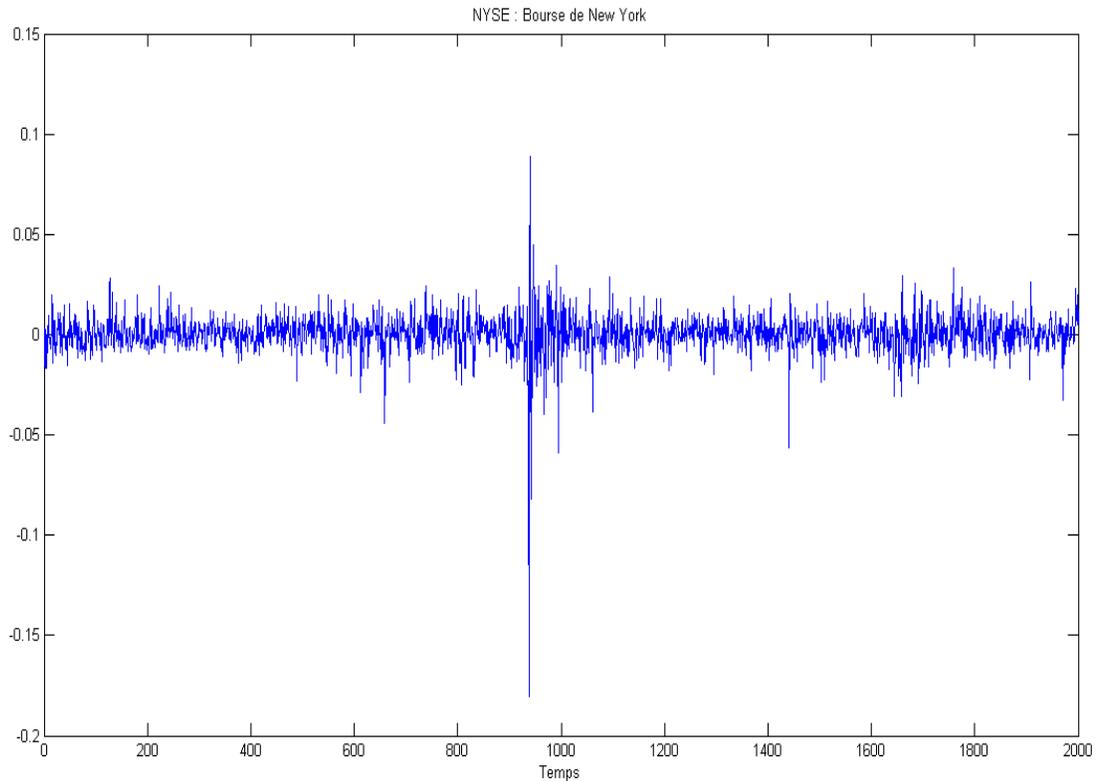


Figure 3.3.1 : Les rendements quotidiens de la Bourse de New York (New York Stock Exchange (NYSE)) à partir du 2 Février 1984 jusqu'au 31 Décembre 1991.

Nous présentons dans ce qui suit, l'estimation des paramètres du modèle ainsi que les graphes de la volatilité obtenus par ces méthodes.

3.3.2 Estimation par la méthode QML

Les paramètres, ω, β, σ , sont estimés par le maximum de vraisemblance et les résultats obtenus pour l'échantillon de 2000 observations sont les suivants :

Paramètres	ω	β	σ
Valeurs initiales	0	0.95	0.2
Estimation	-0.0063	0.9999	0.2048
Ecart type	0.0503	0.029	0.073

La figure montre le journal des rendements au carré, $y_t = \log(r_t^2)$, pour deux cents observations sur les deux mille observations du NYSE ainsi que sa volatilité, en utilisant la méthode QML avec le filtre et lissage de Kalman

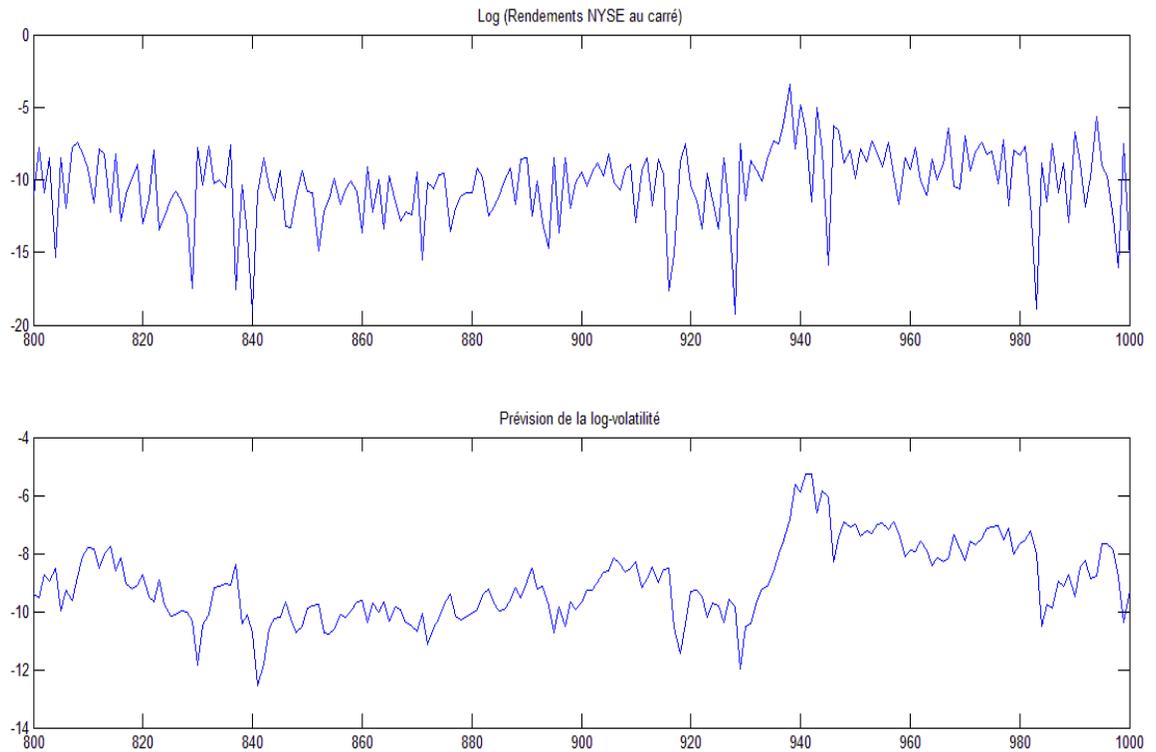


Figure 3.3.2 : Prédiction de la log-volatilité par QML de 800 à 1000.

3.3.3 Estimation par la méthode MCMC

Nous utilisons dans cette partie la méthode de simulation Monte Carlo par les Chaînes de Markov. Cette méthode est appliquée sur le modèle SV, en retenant le même échantillon de la bourse de New York (NYSE) utilisées dans paragraphe précédent. Le résultat de la simulation par la méthode MCMC est résumé dans le tableau suivant :

Paramètres	ω	β	σ
Valeurs initiales	0	0.95	0.2
Estimation	-0.0068	0.963	0.17
Ecart type	0.016	0.007	0.027

Le graphe suivant représente la volatilité extraite à partir de l'échantillon c.à.d. les rendements (NYSE), par la méthode MCMC.

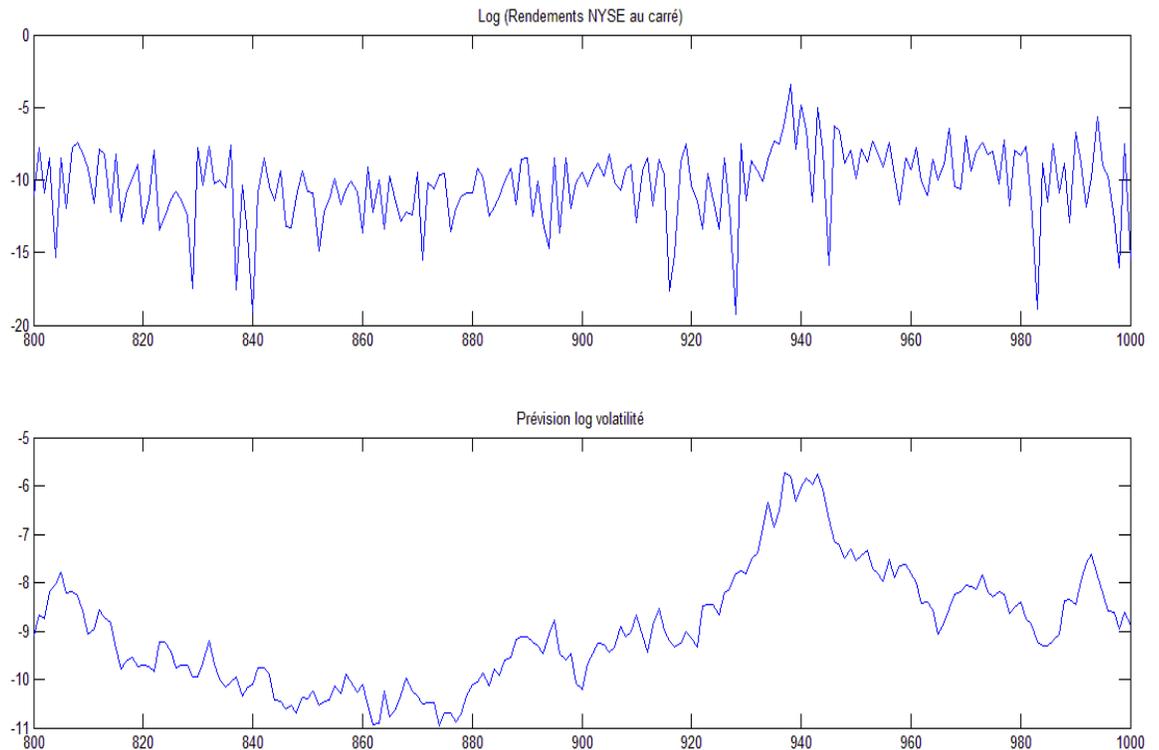


Figure 3.3.3 : Prévion de la log-volatilité par MCMC de 800 à 1000.

3.3.4 Comparaison des QML et MCMC

Afin de comparer la méthode de simulation Monte Carlo par les Chaînes de Markov (MCMC) avec la méthode du quasi maximum de vraisemblance (QML), nous nous basons sur les estimations des paramètres ainsi que les graphes de volatilité lissée.

Les résultats de la simulation montrent clairement que la méthode MCMC est plus pertinente et efficace que la méthode QML par rapport à l'estimation des paramètres.

L'avantage le plus important des procédures MCMC est qu'elles permettent d'obtenir simultanément l'estimation des paramètres, les estimations de la volatilité lissée ainsi que les distributions prédictives des valeurs futures à plusieurs étapes de la volatilité. En plus de cela, si la taille de la simulation est grande alors la méthode MCMC donne asymptotiquement la distribution que la méthode QML. Le principal inconvénient est que la mise en œuvre est assez compliquée et consomme beaucoup de temps par rapport à la méthode QML.

Chapitre 4

Conclusion et Perspectives

Durant cette étude, nous nous sommes intéressés aux modèles de volatilité stochastique à estimer les paramètres du modèle et à capturer la persistance de la volatilité stochastique.

Les approches utilisées pour estimer les paramètres et les volatilités latentes du modèle, sont la méthode MCMC et la QML. L'approche Bayésienne a l'avantage de tenir compte de l'incertitude de ces paramètres, à l'opposé de la méthode du maximum de vraisemblance qui considère ces paramètres comme vraies.

L'inconvénient de la méthode MCMC est qu'elle est compliquée à appliquer contrairement à la méthode QML, mais ces résultats sont plus pertinents que la méthode QML.

Plusieurs axes de recherche sur des modèles plus complexes peuvent être envisagés, tels :

– Les modèles de volatilité stochastique multivariés, dont on cite quelques références :

Yu, J., Meyer, R. (2006). Multivariate stochastic volatility models : Bayesian estimation and model comparison. *Econometric Reviews* 25(2-3) :361-384.

Liesenfeld, R., Richard, J.-F. (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews* 25(2-3) :335-360.

– Les modèles de volatilité stochastique périodiques dont on cite aussi :

Aknouche, A. (2011). On a periodic autoregressive stochastic volatility model. *Computational Statistics and Data Analysis*.

– Les modèles de volatilité stochastique à changement de régime Markovien...

Bibliographie

- [1] Abidoune S. (2009). Volatilité stochastique du taux de change nominal canadien dans une perspective de long et court terme. Université de Montréal.
- [2] Aknouche, A. (2011). On a periodic autoregressive stochastic volatility model. Computational Statistics and Data Analysis. A paraître.
- [3] Billingsley P. (1986). Probability and Measure. 2nd Ed. The University of Chicago.
- [4] Breidt Jay F. et Carriquiry A. L. (1997). Improved Qasi-Maximum Likelihood Estimation for Stochastic Volatility Models. Iowa State University.
- [5] Brockwell P. J. & Richard A. D. (1986). Time Series : Theory and Methods.
- [6] Broto C. & Ruiz E. (2004). Estiamtion methods for stochastic volatility models : A survey. Journal of economic surveys Vol 18, No. 5.
- [7] Chagra D. (2009). Sélection de modèle d'imputation à partir de modèles bayésiens hiérarchiques linéaires multivariés. Université de Montréal.
- [8] Del Moral P. (2009). Ingénierie Stochastique Une introduction avec applications. Centre de recherche INRIA Bordeaux.
- [9] Gendron R. (2005). Analyse bayésienne d'un modèle pour les changements de prix de titres échangés à haute fréquence sur les marchés financiers. Université de Montréal.
- [10] Grewal M. S. & Angus P. A.. Kalman Filtering theory Using MATLAB.
- [11] Halil R. (2011). Inférence Bayesienne en séries chronologiques. Mémoire Magister Université Tizi Ouzou.
- [12] Hamilton J. D. (1954). Time Series Analysis.
- [13] Hayoun M. (2002). La méthode de monte carlo metropolis. Ecole Polytechnique Paris.

- [14] Hedibert F. L.. Markov Chain Monte Carlo and sequential monte carlo methods in Stochastic volatility models. The university of Chicago Booth School of Business.
- [15] Hedibert F. L.. Bayesian Analysis in latent factor and longitudinal models.
- [16] Hevia C. (2008). Maximum Likelihood Estimation of an ARMA(p,q) Model. The world bank DECRG.
- [17] Hurlin C. (2007). Maîtrise d'Economie Appliquée Cours de Tronc Commun. U.F.R. Economie Appliquée.
- [18] Jeongeun K. (2005). Parameter estimation in stochastic volatility models with missing data using particle methods and the EM algorithm.
- [19] Laure E. & Lapeyre B. (2001). Introduction aux Méthodes de Monte-Carlo.
- [20] Lemoine M. & Pelgrin F. (2003). Introduction aux modèles espace d'état et au filtre de Kalman. Revue de l'OFCE 86.
- [21] Lubrano M. (2008). Cours (2008).
- [22] Meyn S.P. & Tweedie R.L. (1993). Markov Chains and Stochastic Stability.
- [23] Millet A.. Cours Master 2ème année : Méthodes de Monte Carlo. Université Paris 7 et Paris 1.
- [24] Milton A. & Irene A. S. (1964). Handbook of Mathematical Functions.
- [25] PELIN Ö. (2004). Analysis of stochastic and non-stochastic volatility models. Thèse Middle East University.
- [26] Ruiz E. (1993). Quasi Maximum Likelihood estimation of stochastic volatility models. Journal of Econometrics 63 (1994) 289-306.
- [27] Ross S. M. (2006). Simulation, Statistical Modeling and Decision Science, Fourth Edition.
- [28] Shumway R. H. & Stoffer D. S. (2011). Time Series Analysis and Its Applications 3rd Ed. Springer Texts in Statistics. RUEY S. TSAY (2002). Analysis of Financial Time Series. Financial econometrics. University of Chicago.
- [29] Straumann D. (2005). Estimation in Conditionally Heteroscedastic Time Series Models. Springer.

- [30] Taylor S. & Karlin H. M. (1975). A first course in stochastic process.
- [31] Von Sachs R. & Van Bellegem S. (2005).Séries chronologiques. 4ème Ed. Université catholique de Louvain Belgique.
- [32] Mergner S. (2009). Applications of state space models in finance. Universitätsverlag Göttingen.