

N d'ordre : 07/2011-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 : **Mathématiques**

Spécialité : **Recherche Opérationnelle : Méthodes Stochastiques**

Par

BENAMARA Nadia

Thème

**Modélisation de la Volatilité Stochastique des Séries
Financières périodiquement corrélées**

Soutenu publiquement le 11/10/2011, devant le jury composé de :

Mme.	GUERBYENNE Hafida	MC A	à l'U.S.T.H.B.	Présidente
M.	BENTARZI Mohamed	Professeur	à l'U.S.T.H.B.	Directeur de thèse.
M.	BERKOUN Youcef	MC A	à l'U.M.M.T.O.	Examinateur.
M.	HAMDI Fayçal	MC B	à l'U.S.T.H.B.	Invité.

Alà barakati Allah

Remerciements

Je remercie Dieu pour la force et la persévérance dont il m'a doté

« La reconnaissance est bien un devoir qu'il faut rendre » jean Jaque ROUSSOU

*A mon Directeur de mémoire Monsieur le Professeur **Mohamed BENTARZI** :*

Je vous remercie pour votre soutien, votre patience et votre intérêt au cours de l'élaboration de ce travail de recherche. La clarté et la richesse de votre savoir, votre esprit critique et vos raisonnements empreints de précision m'ont constamment inspirés. J'ai toujours apprécié votre modestie, vos précieux conseils et votre effort incessant afin de me permettre d'élaborer au mieux ce travail. Veuillez trouver ici, l'expression de ma profonde gratitude et de ma sincère reconnaissance.

*Je tiens à exprimer mes remerciements à Madame **Hafida GUERBYENNE** Maître de Conférence A pour l'honneur qu'elle m'a fait en acceptant de présider le jury de ce mémoire pour ses qualités pédagogiques et ses précieux conseils en Post Graduation.*

*Mes remerciements chaleureux s'adressent également à Monsieur **Youcef BERKOUN** Maître de Conférence A à l'UMMTO et à Monsieur **Fayçal HAMDY** Maître de Conférence B (USTHB) pour l'honneur qu'ils m'ont fait d'accepter de faire partie de mon jury de soutenance.*

*J'ai eu la chance de travailler avec M^{elle} **Leila DJEDDOU**, enseignante chercheuse à l'université de Dely Ibrahim avec qui j'ai beaucoup appris et dont certaines parties de mon mémoire sont tirées de ses travaux. Je voudrais lui témoigner ma profonde gratitude et mes plus vifs remerciements pour son aide précieuse et ses conseils. Je n'oublierais jamais son aide et son soutien, notamment dans la programmation ainsi que la documentation utile qu'elle m'a longuement abreuviée.*

Je tiens aussi à remercier tous ceux qui m'ont aidé de près ou de loin à achever ce travail.

Enfin, il ne me paraît pas possible de citer ici nommément tous ceux avec qui j'ai eu contact durant la préparation de ce mémoire, mais qu'ils ne seront jamais oubliés et qu'ils en soient aussi chaleureusement remerciés.

Dédicaces

Je dédie ce modeste travail qui vient couronner un effort de trois ans

A mes très chers parents pour leurs sacrifices qu'ils ont consentis à mon égard et qui rêvent

de me voir ainsi

A mes très cher(e)s Frères , et Sœurs, qui sont la source de ma fierté et de ma confiance.

En particulier, ma chère soeur Fatma qui m'a toujours encouragé à donner le meilleur de

moi-même

A mes amies

Veillez accepter l'assurance de mon estime et ma reconnaissance.

Table Des Matières

Introduction générale	1
1 Séries financières et les modèles à volatilité	6
1.1 Séries financières	6
1.1.1 Introduction	6
1.1.2 Les principales propriétés des séries financières	7
1.2 Les modèles de la volatilité sur les marchés financiers	13
1.2.1 Introduction	13
1.2.2 La famille de modèles ARCH-GARCH et leurs différentes extensions .	14
1.2.3 Modèle à Volatilité stochastique	21
2 Modèles à Volatilité Stochastique et Filtre de Kalman	23
2.1 Modèle à Volatilité stochastique	23
2.1.1 Introduction	23
2.1.2 Modèle canonique	24
2.1.3 Approximation du $\log(\varepsilon_t^2)$	32
2.1.4 Algorithme EM	32
2.1.5 Simulation	38
2.2 Modèle espace d'état et filtre de Kalman	40
2.2.1 Introduction	40
2.2.2 Modèle espace d'état	41

2.2.3	Algorithme du Filtre de Kalman	42
2.2.4	Simulation	47
3	Estimation des paramètres d'un modèle SV	50
3.1	Introduction	50
3.2	La méthode MCMC	52
3.2.1	Introduction aux méthodes de Monte Carlo par chaînes de Markov	52
3.2.2	Méthodes de Monte Carlo par chaînes de Markov	54
3.2.3	Notions sur les chaînes de Markov	54
3.3.	Quelques algorithmes MCMC	57
3.3.1	L'algorithme de Metropolis-Hastings	58
3.3.2	Algorithme de Metropolis-Hastings composant par composant	60
3.3.3	L'échantillonnage de Gibbs	60
3.4	Estimation des paramètres d'un modèle SV.	64
3.4.1	Estimation.	64
3.4.2	Procédure d'estimation	64
4	Application numérique	71
4.1	Simulation et Estimation d'un modèle SV	71
4.2	Application sur une série de données réelle :	79
4.2.1	Description des données FTSE 100	79
4.2.2	Description des données Nikkei 225	87
4.2.3	Ajustement de la série FTSE100 par le modèle SV	95
4.2.4	Ajustement de la sérieNikkie225 par le modèle SV	97
5	Modèle à volatilité stochastique périodique	101
5.1	Introduction	101

5.2 Processus périodiquement corrélé	102
5.2.1 Historique	102
5.2.2 Processus périodiquement corrélé	104
5.2.3 Fonction d'autocovariance d'un processus périodiquement corrélé	104
5.2.4 Processus périodiquement corrélés et processus multivariés stationnaires	105
5.2.5 Relation entre les autocovariances d'un processus périodiquement cor- rélé et celles processus multivariés stationnaires inhérent	105
5.3Présentation du modèle	106
6 Conclusion	109

0.1 Introduction générale

L'analyse des séries temporelles financières permet d'expliquer les phénomènes des marchés financiers. A cet effet, plusieurs modèles ont été mis en évidence afin d'analyser ces séries et par conséquent permettant d'expliquer au mieux les phénomènes des marchés.

Il est bien connu que les séries financières sont caractérisées par plusieurs faits stylisés en particulier la volatilité. La littérature, pour les travaux liés à l'étude de la volatilité, est très diversifiée. Plusieurs approches ont été utilisées pour traiter ce sujet.

La plus grande partie de ces approches utilise une modélisation basée sur les outils des séries chronologiques pour évaluer et interpréter les données. Ces travaux utilisent, dans la majorité des cas, une version des modèles de type ARCH /GARCH. Ces modèles sont utilisés généralement pour des raisons de simplicité mais ils introduisent un déterminisme entre le rendement et la volatilité.

L'autre alternative est les modèles de la volatilité stochastique qui offrent une plus grande souplesse dans l'ajustement des données.

Le modèle à volatilité stochastique (SV) a été présenté par *Taylor* (1982), et *Tauchen et Pitts* (1983) comme une façon de décrire la volatilité variant dans le temps des rendements des actifs, en 1987 les modèles à volatilité stochastique ont donc été introduits par *Hull et White* (1987) comme réponse aux critiques adressées au modèle fondateur de *Black et Scholes* 1973. En effet le modèle a émergé comme une alternative à la famille *ARCH* (hétéroscédasticité conditionnelle autorégressive) (dans le cas discret), car il est directement relié au type de processus de diffusion utilisé dans la théorie du prix des actifs en finance (voir *Melino et Turnbull*, 1990) et de capturer les principales propriétés empiriques souvent observée dans les séries quotidiennes de rentabilité financière (voir, par exemple, *Carnero, Pena et Ruiz*, 2003).

L'utilisation pratique des modèles à volatilité stochastique nécessite l'estimation

des paramètres du modèle de volatilité. Cette estimation est délicate car la volatilité n'est pas une variable d'état observable.

1. Revue de littérature

L'utilisation de l'outil statistique pour modéliser les données financières a probablement commencé en 1900 avec *Bachelier*. Mais c'est *Markowitz* (1952) qui l'utilise pour introduire la variance en finance, dans son approche "moyenne-variance", appliquée à la sélection de portefeuilles. En 1973, *Black* et *Scholes* formalisent l'utilisation de la volatilité pour l'évaluation des options d'achat et de vente d'actions. Celle-ci y apparaît comme l'écart-type de la loi normale qu'est supposée suivre le rendement relatif d'une action.

Depuis lors, les modèles financiers se sont sophistiqués pour mieux rendre compte des observations empiriques (faits stylisés), comme l'excès de kurtosis, la segmentation de la volatilité, l'asymétrie des réponses aux chocs, la corrélation des puissances, la persistance ... Ils tentent de mieux appréhender cette volatilité. Parmi ces modèles, on peut distinguer ceux qui sont basés sur l'observation, dans lesquels une variable va dépendre de ses observations passées, et ceux qui sont basés sur des facteurs inobservables, dans les quels une variable observée va dépendre de processus stochastiques latents.

Pour les modèles basés sur l'observation, *Engle* propose en 1982 les modèles *ARCH* (AutoRégressifs Conditionnellement Hétéroscédastiques); ils furent étendus aux modèles *GARCH*, *ARMA-GARCH*, *E-GARCH*, *I-GARCH*, *M-GARCH*, *GARCH* multivariés... Ces modèles, sont relativement faciles à estimer analytiquement ou numériquement par l'estimateur du maximum de vraisemblance (EMV).

Pour les modèles basés sur des facteurs inobservables, cette méthode peut parfois encore être utilisable. C'est le cas des modèles linéaires, et des modèles à chaînes de Markov cachées. D'autres outils doivent alors être utilisés : filtre de Kalman, algorithme de Kitagawa-Hamilton, algorithme Espérance Maximisation EM.

La variance conditionnelle du taux de rendement d'un actif financier, qu'on appelle souvent la volatilité, a été capturée pour la première fois, en 1982, par Engle moyennant un modèle de type ARCH. Depuis, plusieurs modèles ont été développés dans le but de généraliser et capturer ce phénomène.

Avant *Engel*, les auteurs utilisent des approximations pour traiter ce genre de problème.

Cependant tous ces modèles qui ont été développés introduisent un déterminisme entre le rendement et cette volatilité alors que celle ci est un processus stochastique latent.

Le modèle à la volatilité stochastique a été proposé par *Jacquier, Polson, Rossi* dans leur article, "Bayesian analysis of stochastic volatility models", (1994). Ces auteurs comparent l'approche ARCH avec celle de type bayésien, et montrent le caractère réaliste des modèles de type SV ainsi la difficulté de connaître la fonction de maximum de vraisemblance pour ce genre de modèle.

Ainsi et afin de détourner ce problème, ces auteurs exploitent l'avantage que présente la méthode de Monte Carlo par Chaîne de Marko (MCMC) qui permet de simuler les distributions a posteriori des paramètres, malgré que le fait qu'on ne connaît pas la forme exacte de la distribution. Ces auteurs, simulent la volatilité d'une seule observation pour un temps donné. Et proposent ensuite des améliorations de ce modèle de base en introduisant des queues épaisses dans la distribution du processus d'innovation ε_t et un effet de levier avec une corrélation entre la moyenne conditionnelle et la variance conditionnelle des erreurs.

Shephard et Pitt (1997) simulent des blocs de log-volatilités en utilisant l'algorithme de Metropolis-Hasting . Ils génèrent une variable aléatoire pour le bloc des log-volatilités, l'accepte ou la rejette avec une probabilité donnée, généralement, par le ratio de Hasting.

Durham dans son article, "Monte Carlo methods for estimating, smoothing and filtering one- and two-factor stochastic volatility models", (2005), décrit une approche qui permet de capturer le log-volatilité, pour un et deux facteurs . En s'appuyant sur le travaux de : *Pitt et Shephard* (1997), *Durbin et Koopman* (1997) , *Durbin et Koopman* (2000), *Sandmann et Koopman* (1998) et *Kim, Shephard, et Chib*, (1998). L'auteur décrit, pour un et deux facteurs, les outils d'une analyse fondée sur la fonction de vraisemblance de la volatilité stochastique appliquée dans le domaine de la finance. L'estimation de la fonction de vraisemblance est effectuée en utilisant la simulation qui intègre une variable auxiliaire non observée (le processus de la volatilité) et en maximisant cette fonction de vraisemblance approximative. Cette approche, communément appelée simulation de l'estimation de maximum de vraisemblance (SMLE), a beaucoup d'affinité avec l'approche Bayésienne et La Monte Carlo par Chaîne de Marko (MCMC), mais elle est vaguement liée à la méthode des moments.

J.McCausland, 2008, propose une méthode de tirage des paramètres et des log-volatilités

à partir de leurs distributions a posteriori et essaie de les améliorer par la suite en introduisant des raffinements. L'avantage de cette méthode est dû à l'importance de ses résultats théoriques et pratiques qu'il proposent. Les outils de simulations sont puissants, de calcul efficace pour un nombre d'observations raisonnable.

2. Présentation du mémoire

L'objectif principal de ce mémoire de Magistère en Mathématiques, Spécialité Recherche Opérationnelle : Option Méthodes Stochastiques, dont l'intitulé est :

Modélisation de la volatilité stochastique des séries financières périodiquement corrélées

est d'étudier et de modéliser la Volatilité Stochastique, nous avons appliqué les méthodes de Monte Carlo par chaînes de Markov (MCMC) , une méthode efficace pour estimer les paramètres du modèle, d'autre part, on présente la méthode du filtre de kalman permettant d'estimer la volatilité stochastique. Nous montrons comment on peut l'utiliser pour prévoir, la volatilité des rendements boursiers. Nous avons effectué des simulations intensives. Elles ont confirmé les performances des estimateurs obtenus.

Pour ce faire, nous progressons comme suit :

Chapitre 1 : il comporte deux parties, fournira en premier lieu, quelques notions de bases des séries financières. Dans la deuxième partie, on présente des plus grandes approches des modèles à volatilité , volatilité déterministe et volatilité stochastique.

Chapitre 2 : L'objectif de ce chapitre est la représentation des modèles SV et quelques propriétés de base, ensuite on se base sur la méthode de Filtre de Kalman appliquée à l'estimation de la volatilité stochastique avec un exemple de données simulées.

Chapitre 3 : Ce chapitre a pour but d'introduire quelques notions de bases liées aux techniques MCMC. Après un bref rappel théorique sur les chaînes de Markov, nous abordons la construction de certains algorithmes MCMC les plus utilisés. En dernier lieu, nous présentons des applications de ces algorithmes aux Modèles à volatilité stochastique.

Chapitre 4 : Ce chapitre présente une étude sur deux séries de données réelles, (FTSE 100) ainsi que les données (Nikkei 225), ainsi nous effectuons des simulations intensives. Elles ont confirmé les performances des estimateurs obtenus.

Chapitre 5 : Dans ce dernier chapitre on présente les modèles à Volatilité Stochastique à coefficients périodiques.

On termine notre travail par une conclusion.

Chapitre 1

Séries financières et les modèles à volatilité

1.1 Séries financières

1.1.1 Introduction

La modélisation des séries financières est un problème complexe. Cette complexité n'est pas seulement due à la grande variété des séries utilisées (prix d'action, taux d'intérêt, taux de change etc.), à l'importance de la fréquence d'observation (seconde, minute, heure, jour, etc.) ou à la disponibilité d'échantillons de très grande taille. Elle tient surtout à l'existence de régularités statistiques " faits stylisés " communes à un très grand nombre de séries financières et difficiles à reproduire artificiellement à partir de modèles stochastiques.

Dans un article paru en 1963, *Mandelbrot* mettait en évidence un ensemble de telles propriétés. Ces régularités empiriques, vérifiées et complétées depuis par de nombreux auteurs, apparaissent plus ou moins nettement en fonction de la fréquence d'observation de la série et de sa nature. Les propriétés que nous présentons ci-dessous, sont valables surtout pour des séries de prix d'action.

1.1.2 Les principales propriétés des séries financières

Les séries de prix d'actif et de rendements présentent généralement un certain nombre de propriétés similaires suivant leur périodicité. Soit p_t le prix d'un actif à la date t et r_t le logarithme du rendement correspondant défini comme suit :

Définition 1.1 :

Le rendement d'un actif financier, à l'instant t , noté r_t , est donné par :

$$r_t = \log\left(\frac{p_t}{p_{t-1}}\right) = \log(p_t) - \log(p_{t-1}) \quad (1.1)$$

Variation relative d'un actif.

Le rendement d'un actif financier est lié à la notion de la variation relative d'un actif :

Définition 1.2 :

La variation relative d'un actif à l'instant t , notée R_t , est donnée par :

$$R_t = \frac{(p_t - p_{t-1})}{p_{t-1}} \quad (1.2)$$

ce qui est équivalent à $R_t + 1 = \frac{p_t}{p_{t-1}}$. A partir des définitions du rendement d'un actif et sa

variation relative on peut facilement établir la relation fonctionnelle suivante :

$$r_t = \log(1 + R_t)$$

Exemple : Trajectoire d'Indice Boursier SP500 du 2/1/1970 au 30/9/2008



Figure 1. Trajectoire de l'Indice Boursier SP500 du 2/1/1970 au 30/9/2008

Propriété 1 (Autocorrélations des carrés des variations de prix)

La série r_t^2 associée aux carrés des rendements présente généralement de fortes autocorrélations tandis que les auto-corrélations de la série r_t sont souvent très faibles (hypothèse de bruit blanc).

L'absence d'auto-corrélation des rendements renvoie à la notion d'efficience. en effet l'hypothèse d'efficience du marché implique que les rendements anticipés d'équilibre corrigés du risque ne sont pas prévisibles.

Quelle que soit la définition retenue, les cours ne peuvent varier entre t et $t+1$ qu'en raison de l'arrivée de "nouvelles" non anticipées. Sous l'hypothèse d'anticipations rationnelles, les erreurs de prévisions définies par $\varepsilon_{t+1} = P_{t+1} - E_t P_{t+1}$, (E_t représente l'espérance conditionnelle) doivent être nulles en moyenne et ne doivent être corrélées avec aucune information de l'ensemble Γ_t d'information disponible à la date t . Cette dernière propriété est appelée *propriété d'orthogonalité*. Or, il est possible de démontrer que si ε_t est auto-corrélé alors la propriété d'orthogonalité n'est pas respectée.

Par **exemple**, supposons que ε_t suit un processus $AR(1)$, $\varepsilon_{t+1} = \rho\varepsilon_t + v_t$ où v_t désigne un bruit blanc. L'erreur de prévision $P_{t+1} - E_t P_{t+1}$ ou *profit non anticipé*, est connu en partie à la date t et par conséquent forme une partie de Γ_t , i.e. $E(\varepsilon_{t+1}/\Gamma_t) = 0$. Autrement

dit, l'erreur de prévision à la date t améliore les prévisions à la période suivante et donc aide à prévoir le cours de la date $t + 1$.

Exemple 2 : Rendements d'Indice Boursier SP500 du 2/1/1970 au 30/9/2008

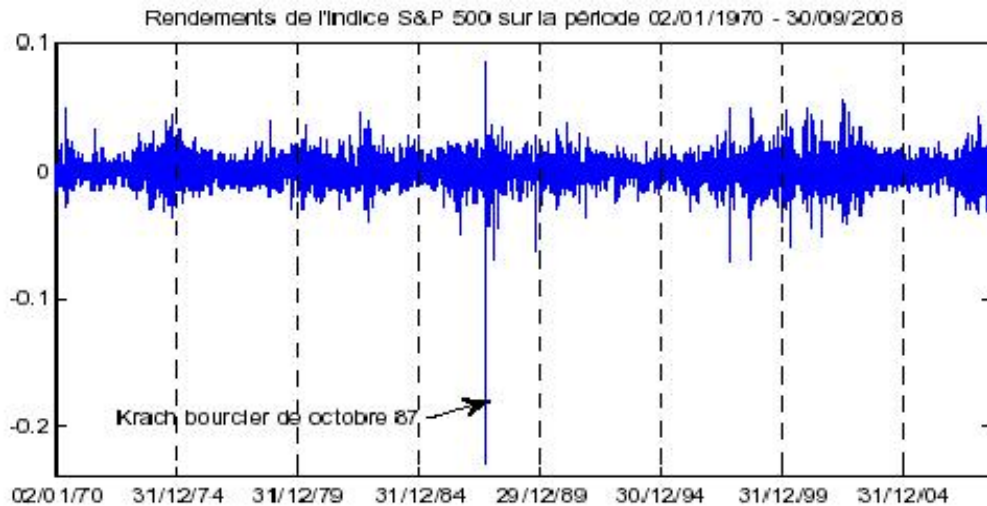


Figure 2. Rendements d'Indice Boursier SP500 du 2/1/1970 au 30/9/2008

Propriété 2 (Distributions à queues épaisses)

L'hypothèse de normalité des rendements est généralement rejetée. Les queues des distributions empiriques des rendements sont généralement plus épaisses que celles d'une loi gaussienne. On parle alors de distribution leptokurtique.

En fait, cet aspect, qui entraîne extrêmement une grande kurtosis (applatissage) relativement à celui de la loi normale fréquemment utilisée, est reconnu depuis les années soixante parmi beaucoup d'investigateurs, en particulier par *Mandelbrot*. D'où la nécessité d'avoir recours à des lois d'innovations dont les queues sont plus épaisses que celles de loi de Gauss.

Définition 1.3 :

On rappelle que la *Kurtosis* d'une variable aléatoire X , de moyenne μ et de variance σ^2 , correspond à son moment centré d'ordre 4.

La *Kurtosis* est une mesure de l'épaisseur des queues de distributions. En règle générale, il est donné par :

$$Kurt = \frac{E(X - \mu)^4}{\sigma^4} \quad (1.3)$$

Cette dernière mesure est comparée à celle de la distribution normale, qui est considérée comme une distribution de référence.

- 1) La distribution pour laquelle $Kurt(X) = 3$ est dite *mésokurtique* (*mesokurtic*).
- 2) Si $Kurt(X) > 3$, alors la distribution de X est plutôt pointue en sa moyenne que celui d'une variable aléatoire normalement distribuée et par conséquent ces queues sont plus épaisses (fat tails) et la loi est dite *leptokurtique* (*leptokurtic*).
- 3) Si $Kurt(X) < 3$, alors la distribution de X est relativement plus aplatie que celle d'une loi normale ces queues sont moins épaisses et la loi est dite *platykurtique* (*platykurtic*).

Le tableau et la figure (3) représente les caractéristiques essentielles de la série SP500 présentée précédemment (la Kurtosis est 5.1876 qui est supérieur de celle de la loi normale 3) :

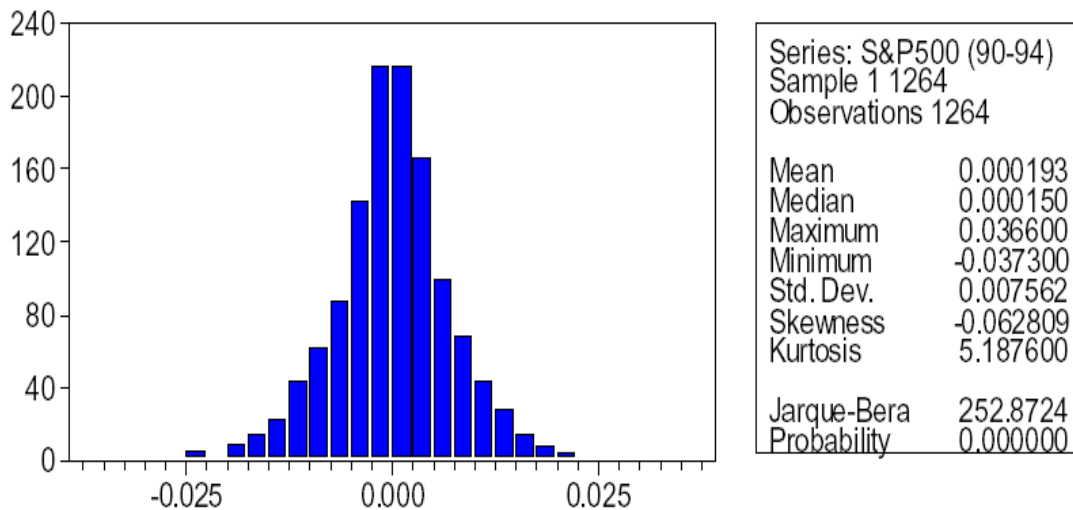


Figure 3. Les caractéristiques de la série d'Indice Boursier SP500

Propriété 3 (Clusters de Volatilité)

Mandelbrot (1963) fait observer une caractéristique importante de la volatilité qu'il résume par une remarque célèbre : "... large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes."

Ce phénomène est connu sous le nom de "**regroupement de la volatilité**" (**volatility clustering**). On observe sans difficulté un tel phénomène sur les rendements du SP500 avec la figure (2). On peut observer que la période de la volatilité faible succède à la période de la volatilité élevée, et vice versa. Une explication possible à ce phénomène a été donnée par *Fama* (1965) l'information aurait tendance à heurter les marchés d'une manière regroupée, ce qui expliquerait un comportement similaire de la volatilité.

Naturellement, ce type de phénomène remet en cause l'hypothèse d'homoscédasticité généralement adopté en économétrie linéaire.

Propriété 4 (Queues épaisses conditionnelles) .

Même une fois corrigée de la volatilité clustering (par exemple avec des modèles ARCH), la distribution des résidus demeure leptokurtique et ce même si la kurtosis est plus faible que dans le cas non conditionnelle.

Propriété 5 (Effet de levier)

Il existe une asymétrie entre l'effet des valeurs passées négatives et l'effet des valeurs passées positives sur la volatilité des cours ou de rendements. On nomme ainsi cette propriété, remarquée par *Brock* (1976), "**Effet de levier**". Les baisses de cours tendent à engendrer une augmentation de la volatilité supérieure à celle induite par une hausse des cours de même ampleur.

Cette propriété n'est pas à confondre avec celle de l'asymétrie de la distribution des cours ou des rendements (**propriété 6**). Il s'agit ici d'une asymétrie de la relation liant les valeurs passés des cours ou rendements à la volatilité de ces derniers.

Propriété 6 (Asymétrie perte/gain)

La distribution des cours est généralement asymétrique : il y a plus de mouvements forts à la baisse qu'à la hausse.

Définition 1.4 :

On rappelle que le coefficient d'asymétrie (Skewness coefficient) d'une variable aléatoire X , sous réserve que son moment d'ordre 3 existe, est donné par :

$$SK = \frac{E(X - \mu)^3}{\sigma^3}. \quad (1.4)$$

1) Si $S_K > 0$, alors la distribution de X est non symétrique. Plus précisément, ceci indique que la probabilité que cette variable prenne des valeurs qui sont supérieures à sa valeur moyenne est plus importante que celle de prendre des valeurs plus petites que la moyenne.

2) Si $S_K < 0$, la distribution est asymétrique et la variable aléatoire a tendance à prendre des valeurs inférieures à sa valeur moyenne avec une probabilité plus grande que celle associée aux valeurs supérieures à la moyenne.

Il est à noter que $S_K = 0$, n'implique pas nécessairement que la distribution est symétrique.

On vérifie ainsi aisément sur *la figure (3)* que l'hypothèse nulle de symétrie de la distribution des rendements sur le *SP500* est rejetée : le coefficient de skewness est significativement négative, dès lors la distribution est non symétrique ; la probabilité d'obtenir des valeurs inférieures à la moyenne étant supérieure à celle d'obtenir des valeurs plus fortes que la moyenne. On retrouve la propriété d'asymétrie aux pertes et gains.

Propriété 7 : (Saisonnalité).

Lorsque les marchés ne fonctionnent pas (week-ends, fêtes), la volatilité tend à augmenter, reflétant ainsi l'information accumulée pendant cet arrêt. Pourtant, on constate souvent que cette hausse est moins forte que si l'information s'accumulait à vitesse constante..

1.2 Les modèles de la volatilité sur les marchés financiers

1.2.1 Introduction

Les rendements des actifs financiers sont qualifiés par leur non constance ; ceci mène à dire que les niveaux de volatilité diffèrent d'un jour à un autre. Pour cette raison, la mesure de la volatilité représente une étape importante dans la détermination de l'évolution des marchés financiers.

L'analyse des séries temporelles financières permet d'expliquer les phénomènes des marchés financiers. A cet effet, plusieurs modèles ont été mis en évidence afin d'analyser ces séries et par conséquent permettent d'expliquer mieux les phénomènes des marchés.

Une littérature abondante a été consacrée au mode de formation de la volatilité des actifs financiers au cours de ces dernières années. Plusieurs approches ont été proposées pour décrire sa dynamique. Toutefois, ce sont les spécifications de type AutoRegressive Conditional Heteroskedasticity (*ARCH*, Engle 1982) ou (*GARCH*, Bollerslev, 1986) qui sont généralement utilisées pour décrire l'évolution de la volatilité des cours boursiers. On peut citer, à titre illustratif, les travaux de French, Schwert et Stambaugh (1987), Chou (1988) ou Baillie et DeGennaro (1990) pour les rendements américains, ou ceux de Poon et Taylor (1992) pour les rendements britanniques. Avec l'accumulation des travaux sur la volatilité, un certain nombre de faits stylisés ont été établis.

Plusieurs études (*Black 1976, Christie 1982, Nelson 1991*) indiquent que réaction de la volatilité à un choc sur le rendement est différent selon le signe du choc : un choc à la baisse a généralement un effet plus fort sur la volatilité qu'un choc à la hausse. Cette propriété est dénommée effet d'asymétrie ou de levier (leverage effect) dans la littérature. De nombreuses spécifications de ces effets ont été proposées pour tester la pertinence de cette hypothèse sur les marchés boursiers (*Nelson, 1991 ; Glosten, Jagannathan et Runkle, 1993 ; Zakoïan, 1994*). Des synthèses de la littérature s'y référant ont également été exposées notamment par *Engle et Ng (1993) et Hentschel (1995)*. En outre, la densité conditionnelle des rendements des actions affiche presque systématiquement un excès de Kurtosis. Certes, l'estimation des modèles GARCH, sous l'hypothèse de normalité des innovations, engendre mécaniquement

un excès de Kurtosis de la distribution non conditionnelle. Toutefois, cette propriété intrinsèque des modèles GARCH n'est pas, en général, suffisante pour rendre pleinement du caractère leptokurtique des séries modélisées.

L'identification de l'excès de Kurtosis traduit donc le fait que la distribution conditionnelle est non-normale. Pour prendre en compte cette propriété, on a recours aux densités conditionnelles autorisant des queues plus épaisses que celle la loi normale. La « distribution de loi de l'erreur généralisée » (*Generalized Error Distribution, ou GED, Nelson 1991*) et la distribution de Student (*Bollerslev, 1987*) sont des distributions permettant de tenir compte de cette caractéristique des rendements boursiers. Des distributions alternatives, fondées sur des développements autour d'une loi de référence (comme les développements en polynômes d'Hermite, proposés par *Gallant et Tauchen, 1989*) ou se présentant sous forme de mélange de lois (loi de poisson et loi normale par exemple), permettent de résoudre, au moins en partie, le problème de non-normalité des distributions conditionnelles des rendements.

Dans cette section nous présentons les plus grandes approches des modèles à volatilité : volatilité déterministe et volatilité stochastique.

1.2.2 La famille de modèles ARCH-GARCH et leurs différentes extensions

Introduction :

Le principe général proposé par *Engle(1982)* consiste à supposer que la variance dépend de l'ensemble informationnel dont on dispose. Il propose une spécification $ARCH(q)$ où le carré des perturbations suit un processus autorégressif d'ordre q . Les modèles $ARCH$ sont donc des modèles autorégressifs conditionnellement hétéroscédastiques. *Engle (1982)* a donc proposer ces processus pour palier aux insuffisances de la classe des représentations $ARMA$, notamment en ce qui concerne les séries financières qui présentent une volatilité (ou variabilité instantanée mesurée par la variance conditionnelle) fonction du temps et par des ajustements asymétriques.

Ainsi, les modèles $ARCH$ sont basés sur une paramétrisation endogène de la variance conditionnelle. La famille des modèles $ARCH$ peut se décomposer en deux sous-ensembles :

les modèles *ARCH* linéaires et les modèles *ARCH* non linéaires.

Les premiers reposent sur une spécification quadratique de la variance conditionnelle des perturbations : modèles *ARCH*(q), *GARCH*(p, q) et *IGARCH*(p, q). Les modèles *ARCH* non linéaires sont caractérisés par des spécifications asymétriques des perturbations. Ce sont les modèles *EGARCH*(p, q), *TARCH*(q) et *TGARCH*(p, q). (Bresson et Pirotte, Séries temporelles)

1. Présentation théorique du modèle ARCH

Les modèles *ARCH* (AutoRegressive Conditional Hétéroskéedasticity), ont été introduit dans le but de mieux analyser les séries financières et monétaires.

Objectivement, le large développement des prévisions, due aux séries chronologiques, provient de l'utilisation de la moyenne conditionnée par le passé du processus. Pour des séries financières, il est évident d'admettre, qu'à afin d'obtenir un meilleur intervalle de prévision, il suffit d'adjoindre des informations supplémentaires du passé à la prédiction de la variance. Le modèle *ARCH* fut la réponse à cette réflexion, et sa conséquence fut immédiate, puisqu'il donna, ainsi, un nouveau souffle à la macro-économie. L'approche la plus utilisée jusqu'alors, pour modéliser l'hétéroscedasticité était d'introduire une variable exogène, afin de prédire la variance de la variable endogène. Cette solution présente le désavantage de requérir à la détermination des causes pouvant induire une variance non constante et de plus, elle ne prend pas en considération le fait que les moyennes et variances conditionnelles puissent évoluer de manière conjointe dans le temps.

En supposant la normalité des innovations, un modèle préférable est donné par Engle (1982), permettant à la variance conditionnelle de changer dans le temps, tout en préservant la variance non conditionnelle constante :

$$\varepsilon_t = \eta_t \sqrt{h_t}, \quad (1.5.a)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2, \quad (1.5.b)$$

$$\text{et } \eta_t \text{ i.i.d. } \rightsquigarrow N(0, 1)$$

Ce modèle a été appelé par Engle (1982) par modèle *ARCH* (q). Il apparaît comme un bruit blanc gaussien, multiplié à chaque temps t par une variable aléatoire dont le carré dépend

de manière linéaire des q dernières valeurs du processus.

Comme on peut l'écrire sous la forme :

$$\varepsilon_t | I_{t-1} \rightsquigarrow N(0, h_t), \quad (1.6.a)$$

$$h_t = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2, \quad (1.6.b)$$

I_{t-1} où désigne l'ensemble d'information disponible à la date $(t-1)$.

Ce qui montre que la distribution conditionnelle de ε_t est normale centrée, de variance h_t , laquelle dépend linéairement des q dernières valeurs du processus

Enfin, en notant , il vient une dernière formulation $u_t = \varepsilon_t^2 - h_t$:

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + u_t, \quad (1.7)$$

Ce qui montre une forme $AR(q)$ pour les carrés du processus d'innovation ε_t^2 , avec toutefois, la remarque que les variables :

$$\varepsilon_t^2 = \eta_t^2 h_t - h_t = (\eta_t^2 - 1) h_t, \quad (1.8)$$

ne sont pas de variance constante mais sont de moyenne nulle et non corrélées entre elles.

Note 1

On note que la dernière formulation donne lieu à des définitions plus générales de modèles *ARCH* où la normalité n'est pas requise.

Note 2

Il faut évidemment que $\alpha_0 > 0$ et que $\alpha_1, \alpha_2, \dots, \alpha_q \geq 0$, sans quoi la variance conditionnelle pourrait être négative.

Bollerslev (1986) a généralisé l'effet *ARCH* en le dotant d'un effet de persistance au niveau de la variance conditionnelle, nous parlons alors de *Generalized ARCH*, ou encore modèle *GARCH*.

2. Présentation théorique du modèle GARCH

Le modèle *GARCH* a été une spectaculaire amélioration du modèle *ARCH*, car en étudiant les différentes applications de ce dernier, il paraît souvent nécessaire d'introduire arbitrairement une structure de retards dans l'équation de la variance conditionnelle pour prendre, en compte la longue mémoire observée dans les travaux empiriques, vu que l'estimation sans la structure de retards imposés, mène souvent à la violation des contraintes de non-négativité [Cf. Engle (1982) ; Engle (1983) ; Engle et Kraft (1983)]. Prenant en considération ces observations, et après de fructueuses discussions, Bollerslev (1986), introduit une dynamique Autorégressive Moyenne Mobile, par analogie avec la démarche usuelle de Box et Jenkins (1976), permettant ainsi, à la structure de retards d'être plus flexible.

L'extension des processus *ARCH* aux processus *GARCH* est similaire à l'extension des processus AR aux processus *ARMA* : elle offre dans la plupart des situations une représentation plus parcimonieuse.

Le modèle *GARCH* (p, q) est donné par :

$$\varepsilon_t | I_{t-1} \rightsquigarrow N(0, h_t), \quad (1.9.a)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \text{ tel que :} \quad (1.9.b)$$

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, 2, \dots, q; j = 1, 2, \dots, p$$

$$\text{et } (q, p) \in \mathbb{N} \times \mathbb{N}$$

Note 3

On peut écrire l'équation (1.9.b) sous la forme symbolique :

$$h_t = \alpha_0 + A(L) \varepsilon_t^2 + B(L) h_t \quad (1.10)$$

où les deux polynômes $A(L)$ et $B(L)$ sont donnés par :

$$A(L) = \sum_{i=1}^q \alpha_i L^i, \quad (1.11.a)$$

$$B(L) = \sum_{j=1}^p \beta_j L^j, \quad (1.11.b)$$

Dans le processus $ARCH(q)$, la variance conditionnelle n'est que fonction du carré des innovations tandis que le processus $GARCH(p, q)$ fait intervenir, en sus, des valeurs passées de la variance conditionnelle, en quelque sorte, il semble correspond à un mécanisme adaptateur d'apprentissage.

3. Présentation théorique des extensions du processus ARCH-GARCH

3.1. Le processus ARMA-GARCH

C'est *Weiss (1986)* qui a introduit dans la variance conditionnelle des effets additionnels des variable expliquées. En effet, la modélisation GARCH peut être appliquée non au processus initial, mais d'innovation. Ceci permet alors d'introduire divers effets additionnels de variables explicatives soit dans la moyenne conditionnelle, soit dans la variance conditionnelle. Par exemple, on peut considérer un modèle de régression linéaire avec erreur GARCH.

$$y_t = x_t b + \varepsilon_t \quad \varepsilon_t \sim GARCH(p, q) \quad (1.12)$$

On peut aussi considérer un modèle ARMA avec erreur GARCH (p, q) :

$$\Phi(L) y_t = \Theta(L) \varepsilon_t \quad \varepsilon_t \sim GARCH(p, q)$$

Enfin, on peut concevoir un modèle ARMA dans lequel la variance non conditionnelle de y peut avoir un effet sur la variance conditionnelle :

$$\begin{aligned} & \Phi(L) y_t = \Theta(L) \varepsilon_t & (1.13.a) \\ \text{et} & E(\varepsilon_t | I_{t-1}) = 0 & (1.13.b) \end{aligned}$$

$$V(\varepsilon_t | I_{t-1}) = c + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \beta_0 [E(y_t / y_{t-1})] + \sum_{j=1}^p \beta_j y_{t-j}^2 \quad (1.14)$$

3.1. Le processus I-GARCH

Les processus $I - GARCH(p, q)$ proposés par *Engle et Bollerslev(1986)*, correspondent au cas d'une racine unitaire dans le processus de variance conditionnelle. Ces modèles sont alors par effet de persistance dans la variance c'est à dire qu'un choc sur la variance conditionnelle actuelle se répercute sur toutes les valeurs futures prévues.

Un processus ε_t satisfait une représentation $I - GARCH$ si et seulement si :

$$V(\varepsilon_t | I_{t-1}) = h_t \quad (1.15.a)$$

$$= \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^p \beta_j \varepsilon_{t-j}^2, \quad (1.15.b)$$

Avec : $\alpha_0 \geq 0$, $\alpha_j \geq 0$ pour $j = 1, \dots, q$ et $\beta_j \geq 0$ pour $j = 1, \dots, p$ tel que :

$$\sum_{j=1}^q \alpha_j + \sum_{j=1}^p \beta_j = 1$$

3.2. Le processus GARCH-M

Engle, Lilien et Robbins(1987) ont proposé un nouveau modèle *GARCH – M* dont la variance conditionnelle représente un variable explicative de la moyenne conditionnelle. C'est un processus adopté à une description de l'influence de la volatilité sur les rendements des titres.

Un processus y_t satisfait une représentation *GARCH – M* si et seulement si

$$y_t = x_t b + \delta h_t + \varepsilon_t \quad (1.16.a)$$

$$= x_t b + \delta V(\varepsilon_t / \varepsilon_{t-1}) + \varepsilon_t \quad (1.16.b)$$

où $\varepsilon_t = z_t \sqrt{\varepsilon_t}$ (1.17.a)

où $z_t \sim N(0, 1)$ et

$$h_t = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^p \beta_j \varepsilon_{t-j}^2, \quad (1.17.b)$$

avec $E(\varepsilon_t | I_{t-1}) = 0$ et $V(\varepsilon_t | I_{t-1}) = h_t$

3.3. Les modèles ARCH/GARCH asymétriques

Diverses études ont montré que la plupart des séries financières sont fortement asymétrique, dans le sens où les variations négatives des prix d'actifs sont suivies par des hausses plus marquées de la volatilité que des variations positives de même ampleurs.

De nombreux modèles se sont développés permettant de détecter cet effet d'asymétrie à

savoir le modèle *GARCH exponentiel* (*EGARCH*) de *Nelson* (1991), le modèle *GARCH à seuil* (*TGARCH*) de *Zakoïan* (1994) le modèle *GARCH asymétrique* (*AGARCH*) de *Engle* et *Ng* (1993).

3.3.1. Le modèle EGARCH

Un inconvénient majeur du modèle GARCH est la symétrie entre les effets des valeurs positives et négatives des innovations sur les rendements des actifs financiers. De plus, il est connu que l’asymétrie est une évidence dans beaucoup de séries chronologiques financières.

Pour surmonter cette faiblesse, *Nelson* (1991) a introduit le modèle GARCH exponentiel (*EGARCH*) pour capturer les effets asymétriques entre les valeurs positives et les valeurs négatives des rendements.

Un processus ε_t satisfait une représentation EGARCH si et seulement si :

$$V(\varepsilon_t | I_{t-1}) = h_t \tag{1.18.a}$$

$$\log(h_t) = \alpha_0 + \sum_{j=1}^q \alpha_j g(z_{t-j}) + \sum_{j=1}^p \beta_j \log(h_{t-j}), \tag{1.18.b}$$

où z_t est une suite de variables aléatoires indépendantes identiquement distribuées gaussiennes centrées réduites et où la fonction $g(\cdot)$ vérifie :

$$g(z_{t-j}) = \theta z_{t-j} + \gamma (|z_{t-j}| - E|z_{t-j}|) \tag{1.19}$$

3.3.2. Le modèle MAR-ARCH

Introduit par *Wong et Li* (2001), pour capturer de l’hétéroscédasticité et une meilleure modélisation des séries temporelles ayant des valeurs aberrantes, des valeurs explosives, des distributions conditionnelles multimodales, et d’autres propriétés de non linéarité et non normalité. Ce type de modèles est composé d’un mélange de K modèles autoregressifs à erreur ARCH, c’est à dire la moyenne conditionnelle suit un mélange de modèles AR et la variance conditionnelle suit un mélange de modèle ARCH.

3.3.3. Les modèles TGARCH et le modèle GJR-GARCH

En 1993, *Glosten, Jaganathan et Runkle* présentent un nouveau modèle non linéaire, le modèle *GJR – GARCH* ou encore *TGARCH*. présenté par *Zakoian* (1994). Sachant que les deux modèles sont similaires sauf que le modèle *TGARCH* spécifie une asymétrie sur l'écart type et non sur la variance conditionnelle.

Outre ces processus, la famille des modèles ARCH-GARCH s'est élargie au cours des années. Pour plus d'approfondissement, on trouve d'autres modèles appartenant a la famille *ARCH* a savoir les modèles *GARCH* à longue mémoire ou nous trouvons les modèles *FIGARCH* qui ont des liens de parenté avec les modèles *IGARCH*, et le modèle *HYGARCH* développé par *Davidson* (2002).

1.2.3 Modèle à Volatilité stochastique

Dans le modèle de la volatilité stochastique, la volatilité est soumise à une source d'erreurs qui peut être liée à celle qui affecte les rendements. Ce modèle se présente sous la forme suivante :

$$y_t = \beta e^{h_t/2} \varepsilon_t, \quad (1.20.a)$$

$$\begin{aligned} h_{t+1} &= \mu + \phi (h_t - \mu) + \sigma_\eta \eta_t, \\ h_1 &\rightarrow N\left(\mu, \frac{\sigma_\eta^2}{1 - \phi^2}\right). \end{aligned} \quad (1.20.b)$$

Où η_t et ε_t sont des bruits blancs Gaussiens qui sont non corrélés centrés et de variance égale à 1.

La modélisation de la volatilité comme une variable stochastique mène immédiatement à des distributions avec des queues épaisses pour les rendements. Le terme autorégressif dans le processus de la volatilité introduit une persistance et une corrélation entre les deux termes innovateurs de ce processus ainsi que le processus du rendement provoquant une asymétrie de la volatilité (*Hull et White (1987)*).

En effet, le terme de bruit de la volatilité rend ce modèle stochastique plus flexible, toutefois, l'inexistence d'une forme rapprochée laisse impossible l'estimation directe via la

méthode du maximum de vraisemblance. L'approche du quasi-maximum de vraisemblance de l'estimation (QMLE) développée par *Harvey, Ruiz et Shephard* (1994).

Les méthodes alternatives sont la méthode des moments généralisés à travers les simulations (GMM) (*Duffie et Singleton*(1993)), ou les solutions analytiques (*Singhton*(2001)), l'approche de vraisemblance à travers l'intégration numérique (*Friedman et Harris*(1988) et l'intégration de Monte Carlo (*Danielsson* (1994), *Pitt et Shephard* (1997), *Durbin et Koopman* (2000)) ou la chaîne de Markov (*Jacquier, Polson et Rossi*(1994), *Kim, Shephard et Chib*(1998)).

Les détails de la description de ce modèle, ses importantes propriétés, l'estimation de ses paramètres sont traités dans les chapitres suivants.

Chapitre 2

Modèles à Volatilité Stochastique et Filtre de Kalman

2.1 Modèle à Volatilité stochastique

2.1.1 Introduction

La classe des modèles de volatilité stochastique est apparue comme une approche alternative pour les modèles de type *ARCH*. Cette approche consiste à formuler un modèle contenant une composante de variance non observable, son logarithme est modélisé directement comme un processus stochastique d'autorégression linéaire. Ces modèles exploitent donc la prévisibilité de la volatilité à partir des variances conditionnelles passées pour déterminer les rendements d'actifs. Une approche plus ancienne explique la dynamique des rendements par le flux d'information (*Clark (1973)*). Cette idée est justifiée par plusieurs études empiriques où l'on observe l'effet de publication de données économiques importantes sur la volatilité (voir, par exemple, *Baillie et Bollerslev [1991]*). Nous pouvons introduire aussi les valeurs absolues des rendements passés pour modéliser une asymétrie dans le comportement de la volatilité suite à un accroissement ou une baisse des prix. Les liens entre la

dynamique des rendements financiers et celles du volume d'échanges peuvent être analysés d'avantage dans le cadre d'un modèle bi-varié de volatilité stochastique.

Nous avons vu que les modèles *ARCH* – *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, y_t , et sa volatilité, h_t , sont liés par l'équation

$$y_t = \beta \sqrt{h_t} \varepsilon_t,$$

où ε_t est un bruit blanc indépendant, généralement supposé indépendant du processus h_t . On complète le modèle en spécifiant la dynamique de h_t , a priori quelconque pourvu qu'elle soit compatible avec la positivité 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 (y_t) et (σ_t) ne sont évidemment pas indépendants mais, contrairement au cas *GARCH*, la variable h_t ne s'interprète plus comme la variance de h_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 Modèle canonique

Le modèle canonique et transformation :

Le modèle à volatilité stochastique canonique semble avoir été introduit dans la littérature économétrique par *Taylor* (1986). Les méthodes statistiques permettant de l'utiliser

en pratique n'ont cependant été développés que dans le courant des années 90. Si les propriétés probabilistes de ces processus sont simples à établir, du moins par comparaison aux modèles ARCH/GARCH, l'inférence statistique pose, comme nous le verrons, de nombreux problèmes pouvant justifier l'utilisation de méthodes sophistiquées.

Nous commençons par passer en revue les propriétés élémentaires du modèle canonique.

Définition 2.1 :

Considérons le modèle introduit par Kim, S, Shephard, N. et Chib, S. (1998), dans l'article Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models, présenté comme suit :

$$y_t = \beta e^{h_t/2} \varepsilon_t, \quad (2.1.a)$$

$$\begin{aligned} h_{t+1} &= \mu + \phi (h_t - \mu) + \sigma_\eta \eta_t, \\ h_1 &\rightarrow N \left(\mu, \frac{\sigma_\eta^2}{1 - \phi^2} \right). \end{aligned} \quad (2.1.b)$$

où η_t et ε_t sont des bruits blancs Gaussiens qui sont non corrélés centrés et de variance égale à 1.

La variable y_t représente le taux de rendement d'un titre financier soit observable et dépend de la variable h_t qui est pas observable. Cette dernière représente le log-volatilité à l'instant t qui est supposée suivre un processus stationnaire ($|\phi| < 1$).

Le modèle (2.1), appelé modèle à volatilité stochastique canonique, a été le plus étudié dans la littérature. L'amplitude de y_t est proportionnelle à h_t mais son signe est indépendant de cette variable. Le log-volatilité h_t étant dépendant de son passé, il est clair que le module de y_t dépend également de celui de ses valeurs passées.

Dans ce modèle, le paramètre ϕ joue le rôle du coefficient de persistance. En effet, que lorsque ϕ est proche de 1, un choc positif sur la volatilité (assimilable à une grande valeur positive de ε_t) a généralement pour effet de maintenir la volatilité à un niveau élevé sur plusieurs périodes. L'effet du choc est évidemment atténué, puisque $\phi < 1$, et tend à disparaître. Un choc négatif a inversement un impact négatif sur la volatilité. Si ϕ est proche de 0, l'effet du choc est transitoire, la volatilité dépendant peu de ses valeurs passées. Enfin,

¹ Kim, Shephard et Chib, (1998)

si ϕ 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 génère le même type d'effets alternés. Ceux-ci n'étant généralement pas observés pour les séries financières. L'interprétation des autres coefficients est plus immédiate. Le paramètre β tel que ($\beta = \exp(\mu/2)$) est un facteur d'échelle pour la volatilité tandis que σ_η mesure l'amplitude des oscillations autour de sa moyenne : la volatilité de la log-volatilité est constante dans ce modèle.

Mais précisément parce que l'on a cet état non-observable (latent), il est nécessaire d'utiliser le Filtre de Kalman pour intégrer l'état inobservable qui sera détaillé dans le prochain chapitre.

Sous l'hypothèse : $E[y_t^4] < \infty$, les deux premiers moments de y_t sont donnés par :

$$E[y_t^2] = E[\exp(h_t)] E[\varepsilon_t^2] = \exp(\mu_h + \frac{\sigma_h^2}{2}) \quad , \quad (2.2)$$

$$E[y_t^4] = E[\exp(2h_t)] E[\varepsilon_t^4] = 3 \exp(2\mu_h + 2\sigma_h^2) \quad . \quad (2.3)$$

Tel que :
$$\begin{cases} \mu_h = \mu \\ \sigma_h^2 = \frac{\sigma^2}{1 - \phi^2} \end{cases}$$

Par conséquent, la Kurtosis est égale :

$$K(y_t) = \frac{E[y_t^4]}{E[y_t^2]^2} = 3 \exp(\sigma_h^2) = 3 \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right) \quad (2.4)$$

La fonction d'autocorrelation de y_t^2 est donnée par :

$$Corr(y_t^2, y_{t-\tau}^2) = \frac{\exp(\sigma_h^2 \phi^\tau) - 1}{3 \exp(\sigma_h^2) - 1} \quad , \quad \tau = 1, 2, \dots, \quad (2.5)$$

et se décroît donc de façon exponentielle dans τ .

²Nikolaus Hautsch et Yangguoyi Ou (2008) , Discrete-Time Stochastic Volatility Models and MCMC-Based Statistical Inference

Par conséquent, pour $\phi \in (0, 1)$, le carré des rendements sont positivement autocorrélées.

Démonstration :

1) $E(y_t^2)$:

Nous avons $E(y_t^2) = E[\exp(h_t)] E(\varepsilon_t^2)$

Sachons que $\varepsilon_t \sim N(0, 1)$ c'est à dire $E(\varepsilon_t^2) = 1$

Calculons $E(\exp(h_t))$:

$$E(\exp(h_t)) = \int_{-\infty}^{+\infty} \exp(h_t) \frac{1}{\sigma_h \sqrt{2\pi}} \exp(0.5(h_t - \mu_h)/\sigma_h^2) dh_t$$

On prend $x = h_t$, alors

$$\begin{aligned} E(\exp(h_t)) &= \int_{-\infty}^{+\infty} \exp(x) \frac{1}{\sigma_h \sqrt{2\pi}} \exp[-(x - \mu_h)^2 / 2\sigma_h^2] dx \\ &= \frac{1}{\sigma_h \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp[x - (x - \mu_h)^2 / 2\sigma_h^2] dx \\ &= \frac{1}{\sigma_h \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(A) dx \end{aligned}$$

Notons $A = x - (x - \mu_h)^2 / 2\sigma_h^2$

$$\begin{aligned} A &= x - [(x^2 + \mu_h^2 - 2x\mu_h) / 2\sigma_h^2] \\ &= \frac{-1}{2\sigma_h^2} [x^2 - 2(\mu_h + \sigma_h^2)x + \mu_h^2] \\ &= \frac{-1}{2\sigma_h^2} [(x - (\mu_h + \sigma_h^2))^2 - (\mu_h + \sigma_h^2)^2 + \mu_h^2] \end{aligned}$$

Par élimination du terme μ_h^2 on aura :

$$\begin{aligned} A &= \frac{-1}{2\sigma_h^2} [(x - (\mu_h + \sigma_h^2))^2 - \sigma_h^2(2\mu_h + \sigma_h^2)] \\ &= \frac{-1}{2\sigma_h^2} [x - (\mu_h + \sigma_h^2)]^2 + \left(\mu_h + \frac{\sigma_h^2}{2}\right) \end{aligned}$$

On remplace le terme A dans $E(\exp(h_t))$

$$\begin{aligned} E(\exp(h_t)) &= \frac{1}{\sigma_h^2 \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[\frac{-1}{2\sigma_h^2} (x - (\mu_h + \sigma_h^2))^2 + \left(\mu_h + \frac{\sigma_h^2}{2}\right)\right] dx \\ &= \frac{1}{\sigma_h^2 \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[\frac{-1}{2\sigma_h^2} (x - (\mu_h + \sigma_h^2))^2\right] \exp\left(\mu_h + \frac{\sigma_h^2}{2}\right) dx \end{aligned}$$

Dans ce cas, $E(\exp(h_t)) = \exp\left(\mu_h + \frac{\sigma_h^2}{2}\right)$ et $E(y_t^2) = \exp\left(\mu_h + \frac{\sigma_h^2}{2}\right)$

2) $E(y_t^4)$:

Nous avons $E(y_t^4) = E[\exp(2h_t)] E(\varepsilon_t^4)$

Sachons que $\varepsilon_t \rightarrow N(0, 1)$ c'est à dire $E(\varepsilon_t^4) = 3$

Calculons $E(\exp(2h_t))$:

$$E(\exp(2h_t)) = \int_{-\infty}^{+\infty} \exp(h_t) \frac{1}{\sigma_h \sqrt{2\pi}} \exp[0.5(h_t - \mu_h)/\sigma_h^2] dh_t$$

On prend $x = h_t$, alors

$$\begin{aligned} E(\exp(2h_t)) &= \int_{-\infty}^{+\infty} \exp(2x) \frac{1}{\sigma_h \sqrt{2\pi}} \exp[-(x - 2\mu_h)^2 / 2\sigma_h^2] dx \\ &= \frac{1}{\sigma_h \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp[2x - (x - 2\mu_h)^2 / 2\sigma_h^2] dx \\ &= \frac{1}{\sigma_h \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(A) dx \end{aligned}$$

Notons $B = 2x - (x - 2\mu_h)^2 / 2\sigma_h^2$

$$B = 2x - [(x^2 + 4\mu_h^2 - 4x\mu_h) / 2\sigma_h^2]$$

$$B = \frac{-1}{2\sigma_h^2} [x^2 - 2(2\mu_h + 2\sigma_h^2)x + 4\mu_h^2]$$

$$B = \frac{-1}{2\sigma_h^2} [(x - (2\mu_h + 2\sigma_h^2))^2 - (2\mu_h + 2\sigma_h^2)^2 + 4\mu_h^2]$$

Par élimination du terme $4\mu_h^2$ on aura :

$$\begin{aligned} B &= \frac{-1}{2\sigma_h^2} [(x - (2\mu_h + 2\sigma_h^2))^2 - 2\sigma_h^2(2\mu_h + 2\sigma_h^2)] \\ &= \frac{-1}{2\sigma_h^2} [x - (2\mu_h + 2\sigma_h^2)]^2 + (2\mu_h + 2\sigma_h^2) \end{aligned}$$

On remplace le terme B dans $E(\exp(2h_t))$

$$\begin{aligned} E(\exp(2h_t)) &= \frac{1}{\sigma_h^2 \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[\frac{-1}{2\sigma_h^2} (x - (2\mu_h + 2\sigma_h^2))^2 + (2\mu_h + 2\sigma_h^2)\right] dx \\ &= \frac{1}{\sigma_h^2 \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[\frac{-1}{2\sigma_h^2} (x - (\mu_h + \sigma_h^2))^2\right] \exp(2\mu_h + 2\sigma_h^2) dx \end{aligned}$$

Dans ce cas, on obtient $E(\exp(2h_t)) = \exp(2\mu_h + 2\sigma_h^2)$

Alors $E(y_t^A) = E(\exp(2h_t)) E(\varepsilon_t^A) = 3 \exp(2\mu_h + 2\sigma_h^2)$

Les propriétés probabilistes du modèle¹

Nous étudions maintenant les propriétés probabilistes du modèle.

1. Stationnarité stricte

Appellons non anticipative toute solution de (2.1) appartenant à la tribu engendrée par $\{(\varepsilon_u, \eta_u) : u \leq t\}$. On a le résultat suivant :

Proposition 2.1 :

Le modèle (2.1) admet, sous la contrainte $|\phi| < 1$, une unique solution strictement stationnaire. Cette solution est non anticipative et donnée par

$$y_t = \exp \left\{ \frac{\mu}{2} + \frac{\sigma_\eta}{2} \sum_{i=0}^{\infty} \phi^i \eta_{t-i} \right\} \varepsilon_t, \quad t \in \mathbb{Z} \quad (2.6)$$

Si $|\phi| > 1$, 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.

2. Stationnarité au second ordre

Proposition 2.2 :

Soit $\alpha_i = E \{ \exp(\sigma_\eta \phi^i \eta_t) \}$, $i \geq 0$. Si $|\phi| < 1$ et $\prod_{i=0}^{\infty} \alpha_i < \infty$, le processus (y_t) défini par (2.6) est un bruit blanc de variance :

$$\text{Var}(y_t) = e^\mu \prod_{i=0}^{\infty} \alpha_i \quad (2.7)$$

■

¹Franco et Zakoïan (2009) Modèles GARCH et à volatilité stochastique

Preuve :

On a, d'après les hypothèses d'indépendance sur les suites ε_t et η_t

$$E(y_t) = E \left\{ \exp \left(\frac{\mu}{2} + \frac{\sigma_\eta}{2} \sum_{i=0}^{\infty} \phi^i \eta_{t-i} \right) \right\} E(\varepsilon_t) = 0, \quad (2.8)$$

$$\begin{aligned} Var(y_t) &= E \left\{ \exp \left(\frac{\mu}{2} + \frac{\sigma_\eta}{2} \sum_{i=0}^{\infty} \phi^i \eta_{t-i} \right) \right\}^2 E(\varepsilon_t^2), \\ &= \exp(\mu) \prod_{i=0}^{\infty} E \left\{ \exp \left(\frac{\sigma_\eta}{2} \phi^i \eta_{t-i} \right) \right\}^2 \end{aligned}$$

$$Var(y_t) = e^\mu \prod_{i=0}^{\infty} \alpha_i \quad (2.9)$$

De plus, pour tout $k > 0$, $Cov(y_t, y_{t-k}) = 0$ qui provient de l'indépendance entre ε_t et η_t .

Remarque : (Francq et Zakoïn (2009))

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

De plus

$$Var(y_t) = \exp \left(\mu + \frac{\sigma_\eta}{2(1-\phi^2)} \right) \quad (2.10)$$

Lorsque η_t suit une loi normale centrée réduite.

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

■

²Francq et Zakoïn (2009) Modèles GARCH et à volatilité stochastique.

Formulation du modèle « espace d'états » (Transformation du modèle)

L'avantage de modèles SV représenté par les équations (2.1) est qu'il peut être transformé en un modèle linéaire en prenant le logarithme des carrés des observations

Remarque :

Par la suite, on suppose $\beta = 1$ (problème d'identification)

Harvey, Ruiz, et Shephard (1994) pose que μ est égal à 0 dans l'équation (2.1) car, habituellement, la moyenne des rendements journaliers et intra-journaliers des actions et des taux d'échange est nulle. De manière à linéariser l'équation (2.1), nous élevons y_t au carré et nous l'exprimons sous forme logarithmique.

On peut donc réécrire l'équation (2.1) comme suit :

$$\log y_t^2 = h_t + \log \varepsilon_t^2, \quad (2.11)$$

Nous pouvons développer davantage l'équation (2.11) puisque nous savons que $\varepsilon_t \sim N(0,1)$, On connaît donc par conséquent la distribution de $\log \varepsilon_t^2$, . Elle correspond à une distribution logarithmique χ_1^2 , dont l'espérance est de $E(\log \varepsilon_t^2) = -1,27$ et la variance de $Var(\log \varepsilon_t^2) = 4.93$.

Finalement, le système d'équations que nous voulons estimer est le suivant :

$$y_t^* = h_t + \log \varepsilon_t^2 \quad (2.12.a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \quad (2.12.b)$$

Les équations (2.12.a) et (2.12.b) sont dans la forme appropriée pour utiliser le Filtre de Kalman qu'on va présenter dans la section suivante. L'équation (2.12.a) est l'équation dite de mesure puisque la variable y_t^* est observée. L'équation (2.12.b) est une équation d'état ou de transition puisque h_{t+1} , la variable d'état, n'est pas observée. Cette dernière variable est simulée par le Filtre de Kalman (détaillé dans la prochaine section).

2.1.3 Approximation du $\log(\varepsilon_t^2)$

Proposition 2.3 :

Dans cette section nous avons présenté une approximation du $\log(\varepsilon_t^2)$, la meilleure approximation est un mélange gaussien :

Récrivons l'équation (2.12.a)

$$y_t^* = h_t + z_t \tag{2.13}$$

Tel que $y_t^* = \log(y_t^2 + c)$ et $z_t = \log \varepsilon_t^2$

$$f(z_t) = \sum_{i=1}^K q_i f_N(z_t | m_i - 1.2704, v_i^2) \tag{2.14}$$

L'équation précédente représente un mélange de K lois gaussiennes univariées avec q_i , les probabilités d'acceptation, $m_i = -1.2704$ les moyens, et v_i^2 les variances. et f_N représente la densité d'une loi normale

Les constantes $\{q_i, m_i, v_i^2\}$ sont sélectionnés de sorte que la densité de $\log(\varepsilon_t^2)$ soit approximativement exacte. La constante c est introduit dans la littérature SV par *Fuller*(1996, p.494–7) pour accroître la robustesse de l'estimateur QML du modèle SV à y_t^2 étant très faible. Tout au long nous allons mettre en $c = 0,001$ (*Kim, S., Shephard, N., et Chib, S.*(1998).).

Il convient de noter que la densité du mélange peut également être écrite en termes du variable s_t tel que :

$$z_t | s_t=i \sim N(m_i - 1.2704, v_i^2) \tag{2.15}$$

$$Pr(s_t = i) = q_i \tag{2.16}$$

Dans le cadre de ce mémoire nous avons choisi l'algorithme EM pour approximer $\log(\varepsilon_t^2)$.

2.1.4 Algorithme EM

L'algorithme espérance-maximisation (Expectation-maximisation algorithm, **EM**), proposé par *Dempster et al.* (1977), est un processus itératif qui utilise la distribution des

données complètes pour calculer les estimateurs du maximum de vraisemblance lorsque les données observées sont incomplètes. Ce procédé se fait en deux étapes. La première est l'étape **E** (pour **E**xpectation), c'est-à-dire l'étape d'espérance., elle consiste à prendre l'espérance conditionnelle de la fonction de log-vraisemblance des données complètes sachant les données observées. La deuxième étape est l'étape **M** (pour **M**aximisation), consiste à trouver l'estimateur qui maximise l'équation trouvée sous l'étape **E**.

Ces étapes sont répétées itérativement jusqu'à convergence

Formulation de l'algorithme EM :

Dans la présentation de *Dempster, Laird et Rubin*, on oppose les données dites incomplètes représentées par la variable aléatoire y de densité $g(y|\phi)$ aux données dites complètes $x = (y', z)'$ formées de la suite des données incomplètes y et des données manquantes z et de densité $f(x|\phi)$. Aux variables aléatoires x et y correspondent respectivement les espaces d'échantillonnage X et Y qui sont liés entre eux par une application de X dans Y .

Comme on n'observe pas $x \in X$, mais seulement $y = y(x) \in Y$, on peut spécifier de façon générale la relation entre les deux types de variables (complètes et incomplètes) par :

$$g(y/\phi) = \int f(x/\phi) dx, \tag{2.17}$$

où X_y est un sous-espace observable de X défini par l'équation $y = y(x)$ (espace dit antécédent de Y), soit $X_y = \{x \in X, y = y(x)\} \subset X$.

Dans son acception générale, l'algorithme EM se définit par les deux phases suivantes.

Phase E dite «Expectation» (ou Espérance)

Sachant la valeur courante du paramètre $\phi^{[t]}$ à l'itération $[t]$, la phase E consiste en la détermination de la fonction :

$$Q(\phi, \phi^{[t]}) = E_c^{[t]} [L(\phi; x)] \tag{2.18}$$

Où $x = (y', z)'$, $Q(\phi; \phi^{[t]})$ l'espérance conditionnelle de la logvraisemblance des données complètes par rapport à la distribution des données manquantes z sachant les données

incomplètes y et la valeur courante $\phi^{[t]}$ du paramètre soit :

$$Q(\phi; \phi^{[t]}) = \int L(\phi; y, z)h(z|y, \phi = \phi^{[t]})dz . \quad (2.19)$$

Avec une spécification générale des données complètes, cette fonction s'écrit

$$Q(\phi; \phi^{[t]}) = \int L(\phi; x)k(x|y, \phi = \phi^{[t]})dx \quad (2.20)$$

où

$$k(x|y, \phi = \phi^{[t]}) = f(x|\phi)/g(y|\phi) \quad (2.21)$$

Phase M dite «Maximisation»

On actualise la valeur courante du paramètre en maximisant la fonction obtenue à la phase E par rapport à ϕ , soit

$$\phi^{[t+1]} = \arg \max_{\Phi} Q(\phi; \phi^{[t]}) . \quad (2.22)$$

$$\phi^{[t+1]} = \arg \max_{\Phi} Q(\phi; \phi^{[t]}) . \quad (2.23)$$

Il existe une version généralisée de l'algorithme dite GEM dans laquelle la valeur actualisée ne maximise pas nécessairement Q mais l'augmente simplement c'est-à-dire, satisfait

$$\left. \begin{aligned} \{ Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \forall t. Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \\ t. Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \forall t. \} \end{aligned} \right\} \quad (2.24)$$

$$\left\{ Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \forall t. Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \right.$$

$$\left. t. Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \forall t. \right\}$$

$$Q(\phi^{[t+1]}; \phi^{[t]}) \geq Q(\phi^{[t]}; \phi^{[t]}), \forall t. \quad (2.25)$$

Cas d'un mélange gaussien :

Dans l'algorithme précédent, la phase de maximisation ne peut être décrite car elle dépend du modèle théorique choisi pour les classes. L'algorithme suivant exprime celle-ci dans le cas d'un mélange de composantes gaussiennes non contraintes.

En pratique, cet algorithme fournit de bons résultats même s'il n'échappe pas aux principaux problèmes des algorithmes de classification :

- 1) L'algorithme EM converge de manière presque sûrement vers un optimum éventuellement local.
- 2) Le résultat final dépend de l'initialisation.
- 3) On doit fixer le nombre de classes cherchées via le nombre de composantes estimées.
- 4) La convergence de l'algorithme peut être très lente.
- 5) L'algorithme peut échouer lorsque le système est mal conditionné (Exemple : inversion de matrices singulières).
- 6) Il est sensible aux données aberrantes

Un exemple particulièrement illustratif des potentialités de l'algorithme EM réside dans son application au cas d'un mélange de distributions (*Dempster et al 1977, Titterington et al 1985, Celeux et Diebolt 1985, McLachlan et Basford 1985, McLachlan et Peel 2000*).

Nous considérerons le cas d'un mélange d'un nombre fixé de lois gaussiennes univariées $N(\mu_j, \sigma_j^2)$ d'espérance μ_j et de variance σ_j^2 en proportion q_j pour chacune des composantes $j = 1, \dots, K$ du mélange.

Soit $y_{N \times 1} = \{y_i\}$ le vecteur des N observations y_i supposées indépendantes et de densité :

$$f_{Y_i}(y; \phi) = \sum_{j=1}^K q_j f_j(y; \theta_j) \quad (2.26)$$

où $q_{K \times 1} = \{q_j\}$, $\phi = (q', \theta'_1, \dots, \theta'_j, \dots, \theta'_K)'$ représentent les paramètres et $f_j(y; \theta_j)$ est la densité de la loi $N(\mu_j, \sigma_j^2)$ relative à la composante j du mélange.

Compte tenu de l'équation (2.26) et de l'indépendance des observations, la logvraisemblance des données observées s'écrit :

$$L(\phi; y) = \sum_{i=1}^N \ln \left(\sum_{j=1}^K q_j f_j(y; \theta_j) \right) \quad (2.27)$$

L est une expression qui ne se prête pas aisément à la maximisation.

Une façon de contourner cette difficulté est d'avoir recours à l'algorithme EM. On introduit alors des variables s_i non observables indiquant l'appartenance de l'observation i à une certaine composante j du mélange telle que $Pr(s_i = j) = q_j$. Par définition, cette appartenance étant exclusive, la densité $g(x_i; \phi)$ du couple $x_i = (y_i, z_i)'$ peut alors s'écrire

$$g(x_i; \phi) = \prod_{j=1}^K [g(y_i, s_i = j; \phi)]^{a_{ij}} \quad (2.28)$$

a_{ij} désignant l'indicatrice $a_{ij} = I_{[s_i=j]}$, soit encore, en décomposant la loi conjointe de y_i et z_i ,

$$g(x_i; \phi) = \prod_{j=1}^K \left(q_j f_j(y_i; \phi_j) \right)^{a_{ij}} \quad (2.29)$$

Les couples x_i étant indépendants entre eux, la densité des données complètes $x = (x'_1, \dots, x'_i, \dots, x'_N)$ est le produit des densités élémentaires soit

$$g(x; \phi) = \prod_{i=1}^N \prod_{j=1}^K [q_j f_j(y_i; \phi_j)]^{a_{ij}} \quad (2.30)$$

On en déduit immédiatement l'expression de la logvraisemblance correspondante

$$L(x; \phi) = \sum_{i=1}^N \sum_{j=1}^K a_{ij} [\ln q_j + \ln f_j(y_i; \phi_j)] \quad (2.31)$$

En prenant l'espérance de $L(\phi; x)$ par rapport à la distribution des données manquantes a_{ij} sachant les données observées et les paramètres pris à leurs valeurs courantes, on obtient :

1) la phase E :

l'expression de la fonction $Q(\phi; \phi^{[t]})$

$$Q(\phi; \phi^{[t]}) = \sum_{i=1}^N \sum_{j=1}^K \alpha_{ij}^{[t]} [\ln q_j + \ln f_j(y_i; \phi_j)] \quad (2.32)$$

où $\alpha_{ij}^{[t]} = E(a_{ij}/y_i, \phi = \phi^{[t]})$ s'interprète comme la probabilité conditionnelle d'appartenance de l'observation i à la composante j du mélange, soit

$$\alpha_{ij}^{[t]} = \Pr(s_i = j/y_i, \phi) = \frac{q_j^{[t]} f_j(y_i; \phi_j^{[t]})}{\sum_{j=1}^K q_j^{[t]} f_j(y_i; \phi_j^{[t]})} \quad (2.33)$$

2) la phase M

Il ne reste plus maintenant qu'à maximiser la fonction $Q(\phi; \phi^{[t]})$ par rapport à ϕ , ou plus précisément

$$Q(\phi; \phi^{[t]})^* = Q(\phi; \phi^{[t]}) - \lambda \left(\sum_{j=1}^K q_j - 1 \right) \quad (2.34)$$

pour prendre en compte, grâce au multiplicateur de Lagrange λ , la relation d'exhaustivité qui lie les probabilités d'appartenance.

Les dérivées partielles s'écrivent pour $j = 1 \dots (K - 1)$:

$$\frac{\partial Q^*}{\partial q_j} = \sum_{i=1}^N \frac{\alpha_{ij}^{[t]}}{q_j} - \lambda, \quad (2.35.a)$$

$$\frac{\partial Q^*}{\partial \mu_j} = \sum_{i=1}^N \alpha_{ij}^{[t]} \frac{(y_i - \mu_j)}{\sigma_j^2}, \quad (2.35.b)$$

$$\frac{\partial Q^*}{\partial \sigma_j^2} = -\frac{1}{2} \left\{ \sum_{i=1}^N \alpha_{ij}^{[t]} \left[\frac{1}{\sigma_j^2} - \frac{(y_i - \mu_j)^2}{\sigma_j^4} \right] \right\} \quad (2.35.c)$$

Par annulation, on obtient les solutions à savoir

Estimation de la probabilité a priori $q_j^{[t+1]}$ de la j-ième classe

$$q_j^{[t+1]} = \frac{\sum_{i=1}^N \alpha_{ij}^{[t]}}{N} \quad (2.36.a)$$

Estimation de la moyenne $\mu_j^{[t+1]}$ de la j-ième classe

$$\mu_j^{[t+1]} = \frac{\sum_{i=1}^N \alpha_{ij}^{[t]} y_i}{\sum_{i=1}^N \alpha_{ij}^{[t]}} \quad (2.36.b)$$

Estimation de la variance $\sigma_j^{2[t+1]}$ de la j-ième classe

$$\sigma_j^{2[t+1]} = \frac{\sum_{i=1}^N \alpha_{ij}^{[t]} (y_i - \mu_j^{[t+1]})^2}{\sum_{i=1}^N \alpha_{ij}^{[t]}} \quad (2.36.c)$$

Si l'on avait fait l'hypothèse de variances homogènes, ($\sigma_j^2 = \sigma^2, \forall j$), la formule (2.36.c) s'écrivant :

$$\sigma^{2[t+1]} = \frac{\sum_{i=1}^N \sum_{j=1}^J \alpha_{ij}^{[t]} (y_i - \mu_j^{[t+1]})^2}{N}$$

2.1.5 Simulation

Il est important de s'attarder sur la distribution de $\log(\varepsilon_t^2)$. Pour établir cette distribution, nous avons généré 1000 nombres aléatoires : $\varepsilon_t \sim N(0, 1)$. et nous avons calculé les ε_t^2 et $\log(\varepsilon_t^2)$. Nous avons appliqué l'algorithme EM cité précédemment pour estimer les probabilités, les écarts types et les moyennes du mélange, avec $K = 7$.

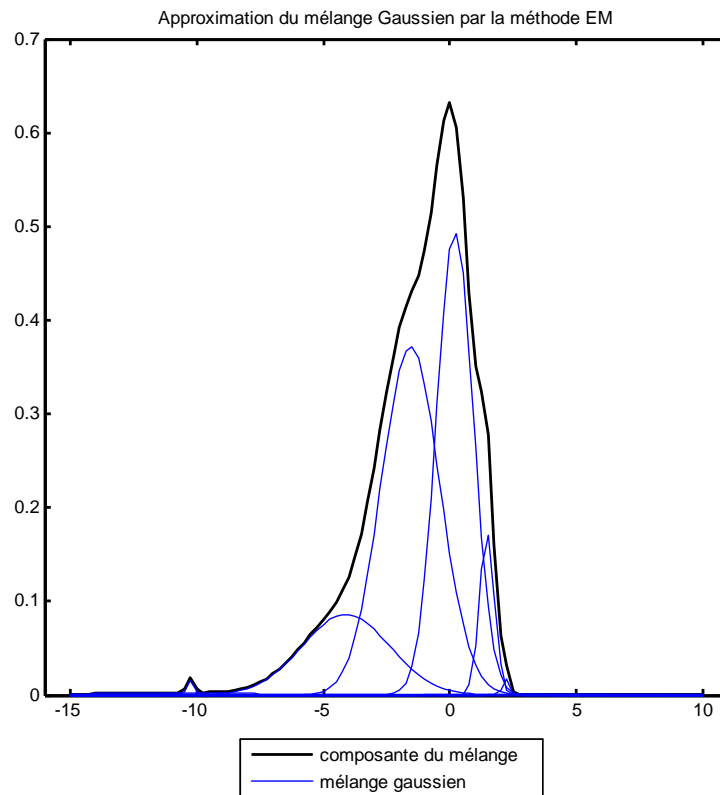
Les résultats d'estimation sont donnés dans le tableau suivant :

s_t	V V(q_i)	$\hat{q}_i = p(s_t = i)$	V V(m_i)	\hat{m}_i	V V(v_i^2)	\hat{v}_i^2
1	0.00930	0.012596	-10.12999	-10.374566	5.79596	5.1341846
2	0.10556	0.1743980	-3.97281	-4.0292851	2.61369	3.5042949
3	0.008	0.0046147	-8.56686	-8.4248846	5.17950	0.0072094
4	0.04395	0.00388411	2.77786	1.93126407	0.16735	0.0001647
5	0.34001	0.3309747	0.61942	0.2972565	0.64009	0.4048415
6	0.04566	0.0306912	1.79518	1.5777450	0.34023	0.1226788
7	0.25750	0.4428408	-1.08819	-1.2739456	1.26261	1.3531950

Selection of the Mixing Distribution to be $\log \chi^2$

V V : Vraie Valeur

La figure suivante illustre un mélange de 7 gaussiennes , Les courbes en blue représentent les densités de probabilité du composante du mélange.



2.2 Modèle espace d'état et filtre de Kalman

2.2.1 Introduction

Le filtre de Kalman est un des outils de traitement du signal dans des domaines très variés, tels que les communications numériques, la géophysique, les disciplines biomédicales et le traitement de la parole. Il s'agit d'une méthode à la fois de prévision et de filtrage, créée par *Kalman (1960)* afin d'estimer l'état d'un système bruité.

Le filtre de Kalman a été utilisé dans un certain nombre d'études pour analyser la variabilité des paramètres dans le temps : dans ce cadre, *Carraro (1986)* a fait une analyse comparative de la méthode de régression et de la méthode du filtre de Kalman ; *Lii-Tarn et al. (2000)* ont fait appel au modèle « espace d'état » ainsi qu'au filtre de Kalman pour étudier l'impact des bulles et de la variabilité des primes de risque sur les prix des titres. Ils ont conclu à la pertinence de la prime de risque dans l'explication des mouvements des titres.

Racicot et Théoret (2005) ont eu recours au filtre de Kalman pour prévoir deux variables financières : la volatilité des taux d'intérêt et de rendements boursiers et le rapport « cours-bénéfice » de l'indice boursier *SP500*.

La forme « espace d'état » et l'algorithme du filtre de Kalman qui lui est associé, permettent donc de tenir compte de la variabilité des paramètres du modèle. Notons en outre que c'est l'un des rares modèles dans lequel la covariance (ou la pente de la régression) peut être présentée sous sa forme conditionnelle en la modélisant soit par un processus AR(1), soit par un processus de marche aléatoire

Le filtre de Kalman procure de nombreux avantages, de par son caractère optimal en tant qu'estimateur linéaire. Cependant, en tant que modèle dynamique, les difficultés sont liées à sa mise en œuvre : la première est celle de l'estimation d'une moyenne qui varie dans le temps.

2.2.2 Modèle espace d'état

Définition 2.2

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

$$y_t = c_t + Z_t \gamma_t + G_t u_t \quad (2.37.a)$$

$$\gamma_{t+1} = d_t + T_t \gamma_t + H_t u_t \quad (2.37.a)$$

Avec $u_t \rightarrow N(0, I)$, les constantes c_t et d_t sont supposées.

La variable y_t est observable et à valeurs dans \mathbb{R}^N tandis que les composantes du vecteur γ_t , appelé vecteur d'état, sont généralement inobservables. Les matrices c_t, Z_t, G_t, T_t, H_t et d_t sont généralement supposées non stochastiques mais peuvent dépendre de t . Les suites $(G_t u_t)$ et $(H_t u_t)$ sont centrées, sériellement non corrélées mais leur variance peut dépendre du temps.

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'est à dire estimer la valeur de γ_t à partir des observations de y jusqu'à la date t ;
- (iii) lisser, c'est à dire estimer la valeur de γ_t à partir des observations de y jusqu'à la date T , avec $T > t$.

Sous des hypothèses de normalité des termes d'erreur et de la distribution initiale du vecteur d'état :

$$\begin{aligned} \gamma_{1/0} &\sim N(a_{1/0}, P_{1/0}) \\ G_t H_t' &= 0, \quad G_t G_t' = \Sigma_t \text{ et } H_t H_t' = \Sigma_{\eta_t} \end{aligned}$$

L'algorithme permet de calculer récursivement la distribution conditionnelle de γ_t sachant y_1, \dots, y_t . Cette distribution est gaussienne et sa moyenne fournit donc un "estimateur" optimal (au sens L^2 (espace de v.a carré intégrable)) de γ_t . Lorsque l'hypothèse de normalité est en défaut, le filtre de Kalman ne fournit généralement plus l'espérance conditionnelle de γ_t . L'estimateur obtenu n'est plus optimal mais seulement optimal parmi les estimateurs linéaires.

Proposition 2.4 (Kim et al (1998))

La forme espace état du modèle (2.1) est obtenue par passage au logarithme :

$$y_t^* = h_t + z_t \quad (2.38.a),$$

$$h_{t+1} = \mu + \phi (h_t - \mu) + \sigma_\eta \eta_t \quad (2.38.b),$$

On note $z_t = \log \varepsilon_t^2$

Et $\varepsilon_t | s_t = i \sim N(m_i, v_i^2)$ et $\eta_t \sim N(0, \sigma_\eta^2)$. ■

On prend $y_t = y_t^*$, $c_t = \mu_i$, $G_t = (\sigma_i \ 0)$, $\gamma_t = h_t$ et $Z_t = 1$. de même $d_t = \mu(1 - \phi)$, $T_t = \phi$ et $H_t = (0 \ \sigma_\eta)$.

Pour la condition de stationarité initiale,

$$a_{1|0} = \mu \text{ et } P_{1|0} = \sigma_\eta^2 / (1 - \phi^2) \ .$$

$$\varepsilon_t = G_t u_t, \quad \eta_t = H_t u_t$$

2.2.3 Algorithme du Filtre de Kalman

Première étape :

Le filtre de Kalman va se dérouler comme suit :

Pour $t = 1 \dots n$:

$$\gamma_{t+1/t} = d_t + T_t \gamma_{t/t-1} + K_t v_t, \quad (2.39)$$

$$P_{t+1/t} = T_t P_{t/t-1} L_t' + \Sigma_{\eta_t}, \quad (2.40)$$

$$F_t = Z_t P_{t/t+1} Z_t' + \Sigma_t, \quad (2.41)$$

$$K_t = T_t P_{t/t-1} Z_t' F_t^{-1}, \quad (2.42)$$

où

$$\gamma_{t+1/t} = E(\gamma_{t+1} / y_1, y_2, \dots, y_t),$$

Dont l'erreur quadratique moyenne associée est notée par $P_{t+1/t}$, et v_t et L_t sont données par :

$$v_t = y_t - Z_t \gamma_{t/t-1} - c_t \quad \text{et} \quad L_t = T_t - K_t Z_t.$$

Deuxième étape :

Le signal de simulation lissage¹ (de Jong et Shephard (1995)) s'inspire de la densité a posteriori normale multivariée

$$(c_1 + Z_1 \gamma_1, \dots, c_n + Z_n \gamma_n) | y, \theta \quad ((2.43))$$

θ représente le vecteur du paramètres, prenant $r_n = 0$ et $N_n = 0$, et

$$D_t = F_t^{-1} + K_t' N_t K_t, \quad (2.44)$$

$$n_t = F_t^{-1} v_t - K_t' r_t. \quad (2.45)$$

Pour $t = n \dots 1$

$$C_t = \Sigma_t - \Sigma_t D_t \Sigma_t, \quad \kappa_t \sim N(0, C_t), \quad (2.46)$$

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, \quad (2.47)$$

$$V_t = \Sigma_t (D_t Z_t - K_t' N_t T_t), \quad (2.48)$$

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t. \quad (2.49)$$

Par la suite, $y_t - \Sigma_t n_t - \kappa_t$ est tiré à partir du signal :

$$c_t + Z_t \gamma_t | y, \theta, c_{t+1} + Z_{t+1} \gamma_{t+1}, \dots, c_n + Z_n \gamma_n \quad (2.50).$$

Démonstration :

1) $\gamma_{t+1/t}$:

En prenant l'espérance de y_t conditionnellement à y_1, y_2, \dots, y_{t-1} on obtient :

$$\begin{aligned} y_{t/t-1} &= E(c_t + Z_t \gamma_t + G_t u_t / y_1, y_2, \dots, y_{t-1}) \\ &= c_t + Z_t E(\gamma_t / y_1, y_2, \dots, y_{t-1}) \\ &= c_t + Z_t \gamma_{t/t-1} \end{aligned}$$

En notant par v_t l'erreur de cette approximation, on a :

¹kim et al (1998)

$$\begin{aligned}
 v_t &= y_t - y_{t/t-1} \\
 &= c_t + Z_t \gamma_t + G_t u_t - (c_t + Z_t \gamma_{t/t-1}) \\
 &= Z_t (\gamma_t - \gamma_{t/t-1}) + G_t u_t
 \end{aligned}$$

D'autre part on a :

$$\begin{aligned}
 E(\gamma_{t+1}/t) &= E(\gamma_{t+1}/t-1) + E(\gamma_{t+1}/v_t) \\
 \gamma_{t+1/t} &= E(d_t + T_t \gamma_t + H_t u_t/t-1) + E(\gamma_{t+1}/v_t) \\
 &= d_t + T_t \gamma_{t/t-1} + E(\gamma_{t+1}/v_t)
 \end{aligned}$$

Considérons l'approximation linéaire :

$$\begin{aligned}
 \gamma_{t+1} &= B v_t + \zeta_t \\
 d_t + T_t \gamma_t + H_t u_t &= B v_t + \zeta_t
 \end{aligned}$$

D'autre part, on a :

$$d_t + T_t \gamma_{t/t-1} = 0$$

Ce qui permet d'avoir

$$\begin{aligned}
 T_t (\gamma_t - \gamma_{t/t-1}) + H_t u_t &= B v_t + \zeta_t \\
 \gamma_{t+1} - \gamma_{t+1/t} &= B v_t + \zeta_t
 \end{aligned}$$

On multipliant à droite par v_t' , on obtient :

$$\begin{aligned}
 B \text{Var}(v_t) &= E \left(T_t (\gamma_t - \gamma_{t/t-1}) + H_t u \right) \left((\gamma_t - \gamma_{t/t-1})' Z_t' + u_t' G_t' \right) \\
 &= T_t E \left((\gamma_t - \gamma_{t/t-1}) (\gamma_t - \gamma_{t/t-1})' \right) Z_t' + H_t E(u_t u_t') G_t' \\
 &= T_t E \left((\gamma_t - \gamma_{t/t-1}) (\gamma_t - \gamma_{t/t-1})' \right) Z_t' + H_t G_t' \text{ car } E(u_t u_t') = I \\
 &= T_t P_{t/t-1} Z_t', \text{ car } H_t G_t' = 0,
 \end{aligned}$$

D'où on obtient :

$$B = (T_t P_{t/t-1} Z_t') (\text{Var}(v_t))^{-1}$$

Calculons $Var(v_t)$:

$$\begin{aligned}
 Var(v_t) &= E \left\{ \left(Z_t (\gamma_t - \gamma_{t/t-1}) + G_t u_t \right) \left((\gamma_t - \gamma_{t/t-1})' Z_t' + u_t' G_t' \right) \right\} \\
 &= Z_t E \left((\gamma_t - \gamma_{t/t-1}) (\gamma_t - \gamma_{t/t-1})' \right) Z_t' + E \left(Z_t (\gamma_t - \gamma_{t/t-1}) u_t' G_t' \right) \\
 &\quad + G_t E \left(u_t (\gamma_t - \gamma_{t/t-1})' \right) Z_t' + G_t E \left(u_t u_t' \right) G_t' \\
 &= Z_t P_{t/t-1} Z_t' + G_t G_t' \text{ car } E \left(u_t u_t' \right) = I
 \end{aligned}$$

Alors, on a :

$$B = \left(T_t P_{t/t-1} Z_t' \right) \left(Z_t P_{t/t-1} Z_t' + G_t G_t' \right)^{-1}$$

Et donc

$$\begin{aligned}
 E(\gamma_t/t - 1) &= B v_t \\
 &= \left(T_t P_{t/t-1} Z_t' \right) \left(Z_t P_{t/t-1} Z_t' + G_t G_t' \right)^{-1} v_t
 \end{aligned}$$

Alors, on a :

$$\begin{aligned}
 \gamma_{t/t-1} &= d_t + T_t \gamma_{t/t-1} + E(\gamma_t/v_t) \\
 &= d_t + T_t \gamma_{t/t-1} + \left(T_t P_{t/t-1} Z_t' \right) \left(Z_t P_{t/t-1} Z_t' + G_t G_t' \right)^{-1} v_t \\
 &= d_t + T_t \gamma_{t/t-1} + \left(T_t P_{t/t-1} Z_t' \right) \left(Z_t P_{t/t-1} Z_t' + \Sigma_t \right)^{-1} v_t \\
 &= d_t + T_t \gamma_{t/t-1} + K_t v_t
 \end{aligned}$$

2) $P_{t+1/t}$:

Finalement, on montre que $P_{t+1/t} = T_t P_{t/t-1} L_t' + \Sigma_{\eta_t}$. En effet, on a :

$$\gamma_{t+1} = d_t + T_t \gamma_t + H_t u_t ,$$

et

$$\gamma_{t+1/t} = d_t + T_t \gamma_{t/t-1} + K_t v_t$$

$$\begin{aligned}
 \gamma_{t+1} - \gamma_{t+1/t} &= (d_t + T_t \gamma_t + H_t u_t) - (d_t + T_t \gamma_{t/t-1} + K_t v_t) \\
 &= T_t (\gamma_t - \gamma_{t/t-1}) - K_t v_t + H_t u_t \\
 &= T_t (\gamma_t - \gamma_{t/t-1}) - K_t [Z_t (\gamma_t - \gamma_{t/t-1}) + G_t u_t] + H_t u_t, \\
 &= (T_t - K_t Z_t) (\gamma_t - \gamma_{t/t-1}) - (K_t G_t - H_t) u_t.
 \end{aligned}$$

Alors, l'erreur quadratique moyenne est donnée par :

$$\begin{aligned}
 P_{t+1/t} &= E \left((\gamma_{t+1} - \gamma_{t+1/t}) (\gamma_{t+1} - \gamma_{t+1/t})' \right) \\
 &= E \left\{ [(T_t - K_t Z_t) (\gamma_{t+1} - \gamma_{t+1/t}) - ((K_t G_t - H_t) u_t)] \times \right. \\
 &\quad \left. [(T_t - K_t Z_t) (\gamma_{t+1} - \gamma_{t+1/t}) - ((K_t G_t - H_t) u_t)]' \right\}, \\
 &= E \left[(T_t - K_t Z_t) (\gamma_{t+1} - \gamma_{t+1/t}) (\gamma_{t+1} - \gamma_{t+1/t})' (T_t - K_t Z_t)' \right] \\
 &\quad + E \left[(K_t G_t - H_t) u_t u_t' (K_t G_t - H_t)' \right] \\
 &\quad - E \left[(T_t - K_t Z_t) (\gamma_{t+1} - \gamma_{t+1/t}) u_t' (K_t G_t - H_t)' \right] \\
 &\quad - E \left[(K_t G_t - H_t) u_t (\gamma_{t+1} - \gamma_{t+1/t})' (K_t G_t - H_t)' u_t' \right].
 \end{aligned}$$

Ce qui donne :

$$\begin{aligned}
 P_{t+1/t} &= (T_t - K_t Z_t) P_{t/t-1} (T_t - K_t Z_t)' + (K_t G_t - H_t) (K_t G_t - H_t)' \\
 &= T_t P_{t/t-1} (T_t - K_t Z_t)' - K_t Z_t P_{t/t-1} (T_t - K_t Z_t)' \\
 &\quad + (K_t G_t - H_t) (K_t G_t - H_t)' \\
 &= T_t P_{t/t-1} (T_t - K_t Z_t)' - K_t Z_t P_{t/t-1} T_t' + K_t Z_t P_{t/t-1} Z_t' K_t' \\
 &\quad + K_t \Sigma_t K_t' + H_t H_t'
 \end{aligned}$$

En remarquant que $Z_t P_{t/t-1} T_t' = F_t K_t'$, $Z_t P_{t/t-1} Z_t' = F_t - \Sigma_t$, $G_t G_t' = \Sigma_t$ et $H_t H_t' = \Sigma_{\eta_t}$ on peut réécrire l'expression précédente sous la forme suivante :

$$\begin{aligned}
 P_{t+1/t} &= T_t P_{t/t-1} (T_t - K_t Z_t)' - K_t F_t K_t' + K_t (F_t - \Sigma_t) K_t' \\
 &\quad + K_t \Sigma_t K_t' + \Sigma_{\eta_t}
 \end{aligned}$$

2.2.4 Simulation

Dans cette partie on s'intéresse à l'estimation de la volatilité stochastique par l'utilisation de filtre de Kalman.

Les processus générateurs de données et leurs trajectoires sont données comme suit :

Modèle (1) :

$$y_t = e^{h_t/2} \varepsilon_t ,$$

$$h_{t+1} = 0.95 (h_t) + 0.15\eta_t$$

$$h_1 \rightarrow N \left(0, \frac{0.15^2}{1 - 0.95^2} \right)$$

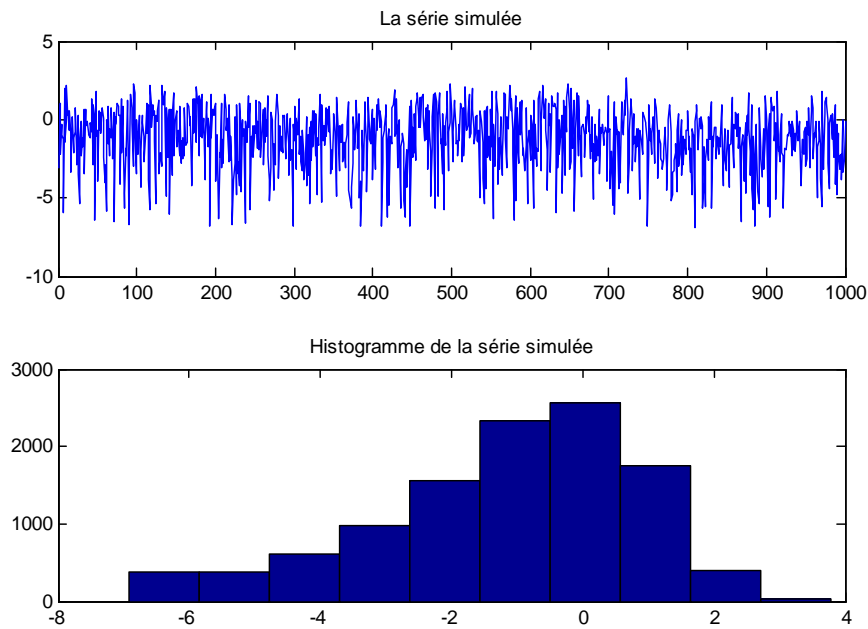


Figure 4. Trajectoire Simulée générée à partir de *Modèle(1)*

Modèle (2) :

$$y_t = 0.6440e^{h_t/2}\varepsilon_t ,$$

$$h_{t+1} = -0.88 + 0.98(h_t + 0.88) + 0.2\eta_t$$

$$h_1 \rightarrow N\left(0, \frac{0.2^2}{1 - 0.98^2}\right)$$

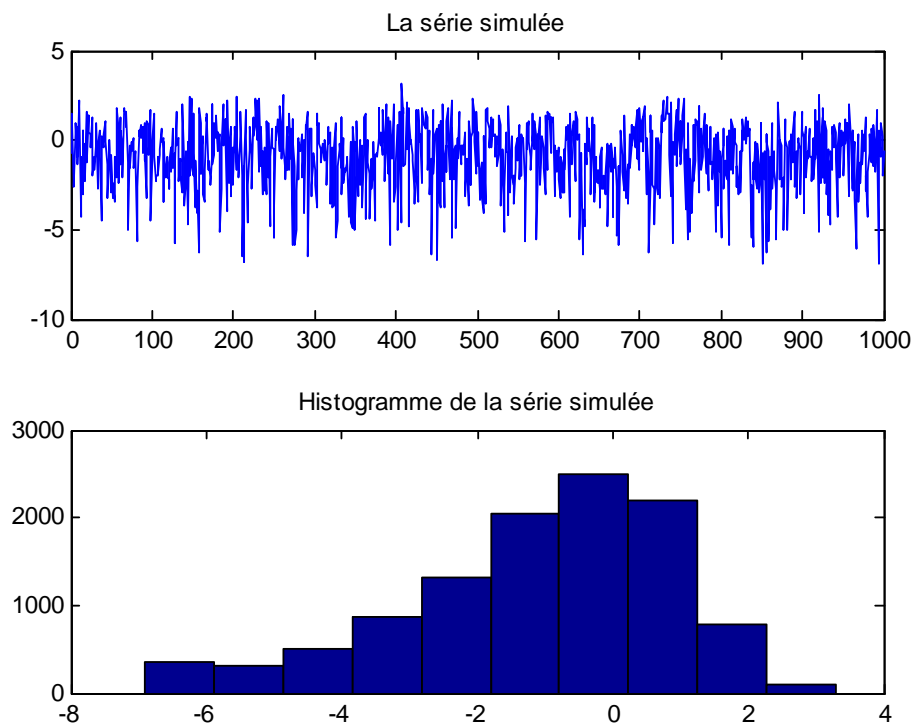
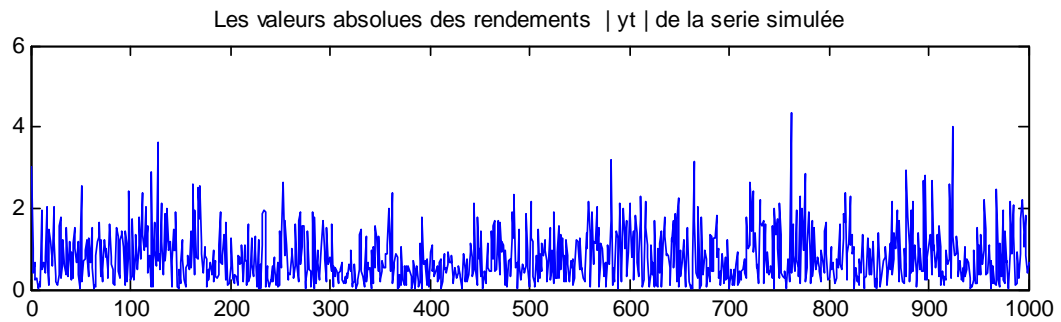
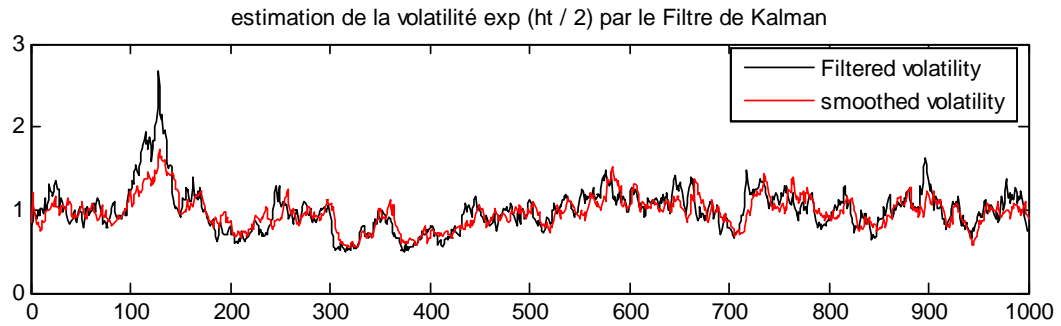


Figure 5. Trajectoire Simulée générée à partir de *Modèle(2)*

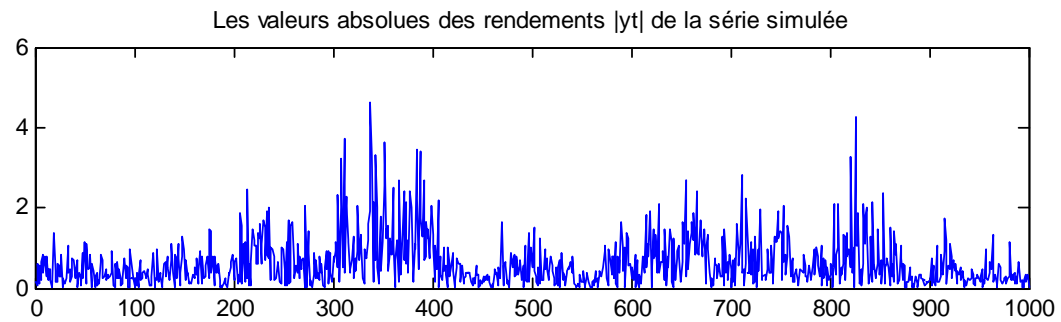
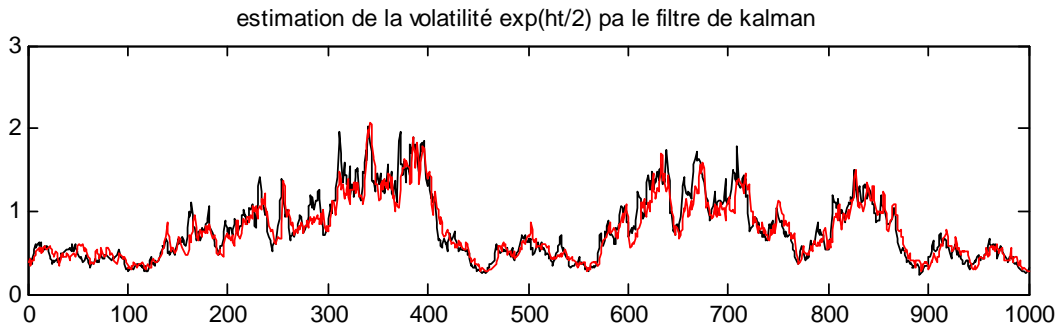
Les résultats des estimations de la volatilité filtrées et lissées sont donnés dans la figure suivante, avec un graphique des valeurs absolues des rendements. Le graphique montre la fonctionnalité attendue de la volatilité filtrée et la volatilité lissée. Tout au long de l'échantillon, la volatilité filtrée est légèrement plus élevée que les valeurs lissées en raison de la baisse progressive de la volatilité observée pour cette série simulée.

Nous avons pris les vraies valeurs des paramètres pour chaque modèle.

Modèle(1)



Modèle(2)



Chapitre 3

Estimation des paramètres d'un modèle SV

3.1 Introduction

Le caractère latent (inobservable) de la volatilité rend l'inférence statistique des modèles à volatilité stochastique plus complexe que celle des modèles de type ARCH-GARCH. En particulier, la loi conditionnelle du processus observé n'est pas explicite, ce qui empêche d'écrire la vraisemblance sous une forme simple permettant de la maximiser. Cette complexité explique que jusqu'au début des années 90, les articles consacrés à l'estimation de tels modèles étaient peu nombreux. Depuis, la situation a radicalement changé et de nombreuses méthodes sont disponibles dans la littérature statistique :

Pseudo-maximum de vraisemblance, moments généralisés, méthode fondée sur les représentations ARMA, méthode Bayésienne et MCMC, utilisée dans ce travail

Dans ce travail, nous avons choisi la méthodologie Bayésienne pour estimer les paramètres du modèle étudié, où chaque paramètre est considéré non plus comme fixe mais comme une variable aléatoire. On peut donc, d'abord choisir des distributions a priori pour les paramètres et par la suite les mettre à jour pour obtenir les distributions a posteriori.

Nous devons (probablement) choisir les distributions a priori de ces paramètres de manière aussi générale que possible pour ne pas influencer les résultats subjectivement. Après cela, on peut considérer la moyenne a posteriori de chacun de ces paramètres pour l'estimation.

Pour certaines distributions les densités a priori et a posteriori appartiennent à la même famille. Dans ce cas, on appelle ces distributions a priori conjuguées.

Pour certains cas élémentaires, les distributions a posteriori sont analytiquement connues et par conséquent on peut calculer la moyenne et la variance des paramètres directement. Cependant en pratique, ceci n'est souvent pas possible et de ce fait on a besoin de techniques numériques de *Chaîne de Markov Monte Carlo (MCMC)*, comme l'Échantillonneur de Gibbs où l'on simule les paramètres et l'état (la volatilité) alternativement.

En outre, on peut rarement simuler directement par la distribution a posteriori, et par conséquent on devra utiliser un algorithme de Metropolis-Hastings (*MH*) où l'on simule par une distribution candidate connue (par exemple la distribution a priori) et ensuite on accepte ou l'on rejette le point simulé.

Ce quotient MH est simplement le quotient de la densité a posteriori et la densité candidate à l'itération k , divisé par la même quantité à l'itération $k-1$. Ce quotient (du moment qu'il est plus petit que 1) peut être interprété comme la probabilité avec laquelle on accepte le point simulé par la distribution candidate. On peut donc comparer cette quantité avec une variable aléatoire uniforme entre zéro et un, et si la première est plus grande que la dernière alors on accepte le point.

Cependant les algorithmes MCMC sont très spécifiques au modèle que l'on choisit. Ils sont donc bien moins flexibles que les algorithmes basés sur la maximisation de la vraisemblance. Nous rappelons les éléments de base de la technique bayésienne et de la méthode MCMC.

3.2 La méthode MCMC

3.2.1 Introduction aux méthodes de Monte Carlo par chaînes de Markov

Les méthodes de Monte Carlo ont été développées pour calculer de manière numérique, et donc approchée, certaines intégrales dont l'expression théorique rendait les calculs rédhibitoires. Depuis que les ordinateurs sont devenus suffisamment puissants, ces méthodes de calcul sont maintenant largement utilisées, et l'on trouve un vaste choix de livres traitant ce sujet. On pourra citer *Gentle, 2002* et *Chen et al. 2000*, ce dernier traitant de manière approfondie les méthodes de Monte Carlo appliquées à l'analyse Bayésienne. Les chaînes de Markov offrent quant à elles un moyen aisé de simuler des valeurs distribuées asymptotiquement selon la loi désirée. La combinaison de ces deux aspects donne lieu à ce que l'on appelle les méthodes de Monte Carlo par chaînes de Markov (MCMC). Il devient alors aisé de calculer certaines intégrales dépendant de lois à partir desquelles on ne peut simuler directement. *Robert, 1996a* et *Robert et Casella, 1999* constituent des ouvrages de référence très complets sur le sujet.

Ce chapitre a pour but d'introduire quelques notions de bases liées aux techniques MCMC. Après un bref rappel théorique sur les chaînes de Markov, nous abordons la construction de certains algorithmes MCMC les plus utilisés. En dernier lieu, nous présentons des applications de ces algorithmes aux Modèles à volatilité stochastique.

Problèmes inhérents à l'approche Bayésienne

Nous introduisons ici quelques unes des difficultés rencontrées par l'approche bayésienne ayant trouvées leurs solutions avec les méthodes MCMC. L'approche bayésienne considère que l'information apportée par les données y est résumée par une loi de probabilité $\pi(\theta|y)$, dite loi a posteriori, qui se déduit de la loi jointe $\pi(\theta)f(y|\theta)$ par calcul de la loi marginale de y notée $m(y)$:

$$\pi(\theta|y) = \frac{\pi(\theta)f(y|\theta)}{m(y)} \text{ avec } m(y) = \int \pi(\theta)f(y|\theta)d\theta ,$$

Où $\pi(\theta)$ est appelée loi a priori sur le paramètre θ . On est alors amené à évaluer des estimateurs d'une fonction de θ , notée $h(\theta)$, sous une certaine fonction de coût $L(\delta, \theta)$, ceci

conduisant à minimiser en δ le coût a posteriori :

$$\int L(\delta, \theta) \pi(\theta|y) d\theta \quad (3.1)$$

Dans le cas particulier d'un coût quadratique

$$L(\delta, \theta) = \|h(\theta) - \delta\|^2, \quad (3.2)$$

L'estimateur de Bayes de la fonction $h(\theta)$ minimisant le coût a posteriori est alors :

$$\begin{aligned} \delta^\pi(y) &= E^\pi [h(\theta)|y] \\ &= \int h(\theta) \pi(\theta|y) d\theta \end{aligned} \quad (3.3)$$

On est ainsi conduit à évaluer une intégrale par rapport à la mesure $\pi(\theta|y)d\theta$. Cet exemple nous permet d'illustrer les problèmes liés à l'utilisation de l'approche bayésienne, problème d'intégration essentiellement. En effet, les difficultés liées au calcul de $\delta^\pi(y)$ sont, d'une part que $\pi(\theta|y)$ n'est généralement pas connue sous forme explicite et, d'autre part, que l'intégration de $h(\theta)$ suivant $\pi(\theta|y)$ ne peut pas être faite analytiquement dans la plupart des cas.

L'inférence statistique traditionnelle se base sur les probabilités d'échantillonnage, $p(y|\theta)$, pour construire le test d'hypothèse et l'intervalle de confiance. Dans le cas d'inférence bayésienne, pour obtenir un estimateur $\hat{\theta}(y)$ de θ , nous avons besoin de sélectionner la caractéristique qui résume la densité a posteriori $f(\theta|y)$, comme la moyenne, le médiane ou le mode. Pour obtenir une mesure de précision d'un estimateur $\hat{\theta}(y)$, on utilise la variance a posteriori de $\hat{\theta}(y)$:

$$var \hat{\theta}(y) = E(\theta|y(\theta - \hat{\theta}(y))^2) \quad (3.4)$$

On montre dans la littérature, Carlin et Louis (2000, chapitre 2), que la moyenne a posteriori est le meilleur estimateur qu'on peut prendre

Tout ceci permet donc de mettre en évidence les problèmes d'intégration comme faisant partie des principaux écueils de la statistique bayésienne, un autre étant constitué par les problèmes de minimisation.

C'est cette classe de problèmes que se proposent d'aborder les méthodes MCMC.

3.2.2 Méthodes de Monte Carlo par chaînes de Markov

Nous allons montrer qu'il est possible d'obtenir un échantillon y_1, \dots, y_m de loi f , sans simuler directement suivant f .

A ce stade de l'étude des méthodes MCMC, il est nécessaire de rappeler quelques résultats et définitions propres aux chaînes de Markov dont nous aurons besoin par la suite.

3.2.3 Notions sur les chaînes de Markov

Une chaîne de Markov homogène à temps discret et d'espace d'état E continu est définie par une séquence de variables aléatoire $\{X^{(0)}, X^{(1)}, \dots\}$ avec $X^{(t)} \in E$, vérifiant la propriété de Markov en temps. Celle-ci signifie que, étant donné l'état courant $X^{(t)} (t \geq 0)$, la loi de $X^{(t+1)}$ est indépendante du passé de la chaîne $\{X^{(0)}, X^{(1)}, \dots, X^{(t-1)}\}$, l'homogénéité signifiant l'invariance de cette propriété par rapport au temps.

Remarque :

L'espace d'état peut être très complexe et contenir aussi bien des espaces discrets que des espaces continus.

La loi d'une chaîne de Markov homogène en temps (discret) $\{X^{(t)}\}$ est alors caractérisée par l'état initial $X^{(0)}$ et par le noyau de transition

$$P(x, A) = P(X^{(t+1)} \in A | X^{(t)} = x) \quad \text{pour} \quad A \subset E . \quad (3.5)$$

Définition 3.1

Un noyau de transition est une fonction P définie sur $E \times B(E)$ ($B(E)$ désignant la tribu Borélienne) à valeurs dans $[0, 1]$ telle que :

1. $\forall x \in E, \quad P(x, \cdot)$ est une mesure de probabilité,

2. $\forall A \in B(E), \quad P(\cdot, A)$ est mesurable.

Ce n'est rien d'autre qu'une probabilité sur $B(E)$, dépendant mesurablement du paramètre x . Lorsque l'espace E est discret, le noyau de transition est simplement une matrice de transition ayant pour éléments $P_{xy} = P(X^{(t+1)} = y | X^{(t)} = x)$ pour $x, y \in E$. Dans le cas continu avec densité, le noyau représente la densité conditionnelle $P(x, x')$ de la transition. C'est-à-dire que

$$P(X \in A | x) = \int_A P(x, x') dx' \quad (3.6)$$

Remarque :

On ne s'intéresse ici qu'aux chaînes de Markov homogènes et n'ayant pas de comportement périodique.

Si $X^{(t)}$ est distribuée selon ν sur E , alors $X^{(t+1)}$ est distribuée selon νP donnée par

$$\nu P(A) = \int P(x, A) \nu(dx) . \quad (3.7)$$

Une chaîne de Markov est de mesure invariante π si $\pi = \pi P$ c'est-à-dire :

$$X^{(t)} \sim \pi \implies X^{(t+1)} \sim \pi . \quad (3.8)$$

Il nous faut déterminer les conditions nécessaires pour pouvoir obtenir un échantillon distribué selon la loi π , ceci dans le but d'approcher des intégrales par rapport à cette loi. On voit facilement qu'une propriété nécessaire est que la chaîne $X^{(t)}$ générée puisse atteindre tous les points du support de π . C'est ce que l'on appelle précisément la propriété d'irréductibilité, indiquant qu'il est alors possible pour la chaîne de joindre deux états quelconques avec une probabilité non nulle.

Définition :

Une chaîne de Markov est dite irréductible (ou φ - irréductible) s'il existe une densité φ sur E telle que pour tout $A \subset E$ avec $\varphi(A) > 0$

$$P(\exists t > 0 \quad X^{(t+1)} \in A \setminus X^{(0)} = x) > 0 \quad \text{pour tout } x \in E . \quad (3.9)$$

Lorsqu'il existe une mesure de probabilité invariante pour une chaîne irréductible, la chaîne est dite positive. Dans un cadre discret, la récurrence d'un état se définit comme une

garantie de retour, c'est-à-dire que la probabilité d'y revenir est égale à un. On devine donc que cette notion est automatiquement satisfaite pour les chaînes irréductibles sur un espace d'état fini. On dit alors que la chaîne $X^{(t)}$ est récurrente et irréductible. Lorsque l'espérance du temps de retour à l'état x est finie, on dit que cet état est récurrent positif.

Lorsque la chaîne est à espace d'état continu, il faut utiliser à la place une notion sensiblement plus forte appelée récurrence au sens de Harris dont on pourra trouver les détails dans *Robert [1996a]*. Si l'on ajoute à cette propriété le fait que la chaîne est positive et aperiodique, on obtient alors ce qui est appelé une chaîne ergodique, cette notion assurant la convergence de la chaîne vers une loi limite.

Nous allons maintenant introduire le concept de chaîne de Markov réversible, définition que l'on peut résumer par la condition suivante

$$\int_A P(x, B)\pi(x)dx = \int_B P(x, A)\pi(x)dx , \quad (3.10)$$

pour tout $x \in E$ et tout A et B inclus dans E .

cette condition signifie que le taux auquel la chaîne se déplace de x à y (i.e. $\pi(x)P(x, y)$) lorsqu'elle est à l'équilibre (lorsque la convergence vers la distribution stationnaire est atteinte), est égal au taux auquel elle se déplace de y à x (i.e. $\pi(y)P(y, x)$). La condition ci-dessus est souvent appelée *condition de balance*. Les chaînes de Markov vérifiant cette condition sont appelées des chaînes réversibles.

La réversibilité permet de montrer des résultats de type théorème de la limite centrale (de *Robert, 1996a*). Les conditions nécessaires à ce type de théorème sont assez contraignantes, mais la réversibilité permet de les obtenir. L'intérêt est que la réversibilité, généralement très contraignante, est facile à imposer dans la plupart des algorithmes de Monte Carlo par chaînes de Markov grâce à des étapes de simulation supplémentaires (cf *Green [1995]* ou *Tierney [1994]*).

Les chaînes réversibles sont très utiles car pour toute chaîne irréductible de noyau de transition P , si π est une distribution vérifiant (3.10), alors la chaîne est récurrente au sens de Harris, réversible et de loi stationnaire π .

De manière intuitive, on peut considérer qu'une chaîne de Markov récurrente irréductible de loi invariante π va nous permettre d'approcher les intégrales de la forme $\int_E h(y)f(y)dy$. C'est ce que précise le théorème suivant (*Tierney, 1996, p65*) :

Théorème :

Soit $\{X^{(0)}, \dots, X^{(t-1)}\}$ une chaîne de Markov irréductible sur E de noyau de transition P et de distribution invariante π . Soit h une fonction à valeurs réelles sur E π -intégrable.

Alors

$$P \left(\frac{1}{m} \sum_{j=1}^m h(x_j) \longrightarrow \int_E h(x)\pi(x)dx \mid X^{(0)} = x^{(0)} \right) = 1 \quad (3.11)$$

Pour presque tout $x^{(0)}$. ■

Ce théorème implique notamment que sous certaines conditions de régularité pour h , la moyenne empirique de l'échantillon généré par une chaîne de Markov de distribution invariante $p(x|y^{(n)})$ converge presque sûrement vers $E(h(X)|y^{(n)})$.

3.3 Quelques algorithmes MCMC

Le principe fondamental des méthodes MCMC consiste donc à utiliser une chaîne de Markov irréductible vérifiant le théorème précédent. Partant d'une valeur initiale arbitraire $x^{(0)} \in E$, on génère une chaîne $(x^{(t)})$ à partir d'un noyau de transition de loi stationnaire f , qui garantit de plus la convergence en loi vers f . Pour T "assez grand", on peut considérer $x^{(T)}$ comme étant distribué suivant f et obtenir ainsi un échantillon $(x^{(T)}, x^{(T+1)} \dots)$ qui est effectivement distribué suivant f , même si les $x^{(T+t)}$ ne sont pas indépendants.

Définition 4.2 :

On appelle algorithme MCMC toute méthode produisant une chaîne de Markov $(x^{(t)})$ ergodique dont la loi stationnaire est la distribution d'intérêt.

Nous allons aborder deux grands types de méthodes MCMC : les méthodes de Metropolis Hastings, très générales, et l'échantillonnage de Gibbs. Ce dernier est d'une importance pratique et historique considérable.

3.3.1 L'algorithme de Metropolis-Hastings

Il repose sur l'utilisation d'une densité conditionnelle $q(y|x)$ par rapport à la mesure dominante pour le modèle, et ne peut être mis en pratique que si $q(\cdot|x)$ est soit simulable rapidement, soit disponible analytiquement. L'algorithme de Metropolis-Hastings associé à la loi objectif f et la loi conditionnelle q produit une chaîne de Markov $(x^{(t)})$ générée par l'algorithme suivant :

Algorithme :

Etant donné $x^{(t)}$

1- Générer $y_t \sim q(\cdot|x^{(t)})$

2- Générer une v.a B à partir d'une loi du Bernolli de paramètre $\rho(x^{(t)}, y_t)$

où
$$\rho(x^{(t)}, y_t) = \min \left[1, \frac{f(y_t) q(x^{(t)}/(y_t))}{f(x^{(t)}) q(y_t/x^{(t)})} \right]$$

3- Prendre

$$x^{(t+1)} = \begin{cases} y_t & \text{si } B = 1 \\ x^{(t)} & \text{si } B = 0 \end{cases}$$

Remarque : q est appelée la loi instrumentale

Il est nécessaire d'imposer des conditions minimales sur loi conditionnelle q pour que f soit effectivement la loi limite de la chaîne $(x^{(t)})$ produite. Par exemple, si E , support de f , que nous supposons connexe, est systématiquement tronqué par q , c'est-à-dire s'il existe $A \subset E$ tel que

$$\int_A f(x)dx > 0 \text{ et } \int_A q(y|x)dy = 0 \quad \forall x \in E, \quad (3.12)$$

L'algorithme de Metropolis-Hastings ne peut admettre f comme loi stationnaire puisque, partant de $x^{(0)} \notin A$, la chaîne $(x^{(t)})$ ne visite jamais A . La condition nécessaire et minimale à imposer est alors :

$$\bigcup_{x \in \text{supp } f} \text{supp } q(\cdot/x) \supset \text{supp } f \quad (3.13)$$

En supposant de plus que $\text{supp } f$ est connexe.

Sous la condition précédente, la chaîne de Markov définie par l'algorithme de Metropolis Hastings admet le noyau de transition suivant :

$$P(x, A) = 1_{[x \notin A]} \int_A q(y|x) \rho(x, y) dy + 1_{[x \in A]} \left[1 - \int_A q(y|x) \rho(x, y) dy \right] \quad (3.14)$$

Ceci représente la probabilité de se déplacer vers l'ensemble $A \subset E$ en partant de n'importe quel x ; le premier terme étant la probabilité d'arriver dans A en partant de $x \notin A$, le second, la probabilité de rester dans A alors qu'on y était déjà. Ce noyau de transition définit alors une chaîne réversible ayant donc f pour loi stationnaire, qu'on peut résumer par le théorème suivant (*Saint Pierre* 2003)

Théorème :

Pour toute loi conditionnelle q , vérifiant la condition minimale, f est une loi stationnaire de la chaîne $(x^{(t)})$ produite par l'algorithme de Metropolis-Hastings .■

Ce théorème nous donne une mesure de l'universalité de ce type d'algorithmes. Afin de vérifier que celui présenté entre bien dans le cadre des algorithmes MCMC, il faut établir l'ergodicité de $(x^{(t)})$. En se référant à *Tierney, 1994* et *Robert, 1996a*, on apprend qu'il suffit de montrer la f -irréductibilité et l'apériodicité de $(x^{(t)})$.

L'irréductibilité de $(x^{(t)})$ découle de conditions suffisantes comme la positivité : $q(y|x) > 0$ pour tout $(x, y) \in E^2$.

L'apériodicité semble évidente puisque l'algorithme autorise formellement les événements $\{x^{(t+1)} = x^{(t)}\}$. Il faut cependant que la probabilité de ces événements soit non nulle, c'est-à-dire :

$$P [f(x^{(t)}) q(y_t/x^{(t)}) = f(y_t) q(x^{(t)}/y_t)] < 1 \quad (3.15)$$

Cette condition signifie que q ne doit pas être le noyau de transition d'une chaîne de Markov réversible admettant f comme loi stationnaire.

3.3.2 Algorithme de Metropolis-Hastings composant par composant

Les algorithmes habituels mettent à jour tout les composants de $x^{(t)}$ à la fois. Il peut parfois être utile de mettre à jour les différents composants de $(x_1^{(t)}, \dots, x_d^{(t)})$ séparément, c'est ce que fait l'algorithme présenté ici. Les composants seront mis à jour selon leur ordre d'indiciage, c'est-à-dire que l'on mettra successivement à jour $x_1^{(t)}, x_2^{(t)}, \dots, \text{et } x_d^{(t)}$. On notera les lois conditionnelles $f_i(x_i^{(t)}) = f_i(x_i^{(t)}/x_{-i}^{(t)})$ avec

$$x_{-i}^{(t)} = (x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t)}, \dots, x_d^{(t)}).$$

La valeur y_i utilisée pour la mise à jour de $x_i^{(t)}$ est générée selon la densité notée $g_i(y_i/x_i^{(t)})$ même si elle peut dépendre de $x_{-i}^{(t)}$. L'algorithme de mise à jour du $i^{\text{ème}}$ composant à l'étape t est alors :

1- Générer $y_i \sim g_i(\cdot | x_i^{(t)})$

2- Générer une v.a B à partir d'une loi du Bernolli de paramètre ρ_i

$$\text{où } \rho_i = \min \left[1, \frac{f_i(y_i)}{f_i(x_i^{(t)})} \frac{g_i(y_i^{(t)}/(y_t))}{g_i(x_i^{(t)}/y_i)} \right]$$

3- Prendre

$$x_i^{(t+1)} = \begin{cases} y_i & \text{si } B = 1 \\ x_i^{(t)} & \text{si } B = 0 \end{cases}$$

En fait, c'est sous cette forme qu'était présenté l'algorithme dans l'article original de *Metropolis et al* 1953. C'est un compromis entre l'algorithme de Metropolis Hastings présenté plus haut et celui de l'échantillonnage de Gibbs que nous allons voir.

3.3.3 L'échantillonnage de Gibbs

Formellement, celui-ci peut se décrire comme un cas particulier de l'algorithme de Metropolis Hastings, cependant il s'en distingue par plusieurs caractéristiques.

1. Le taux d'acceptation est uniformément égal à 1 (Toutes les valeurs simulées sont acceptées).

2. Son utilisation entraîne des limitations fortes sur le choix des paramètres des lois instrumentales et suppose la connaissance de certaines propriétés de f .

3. L'échantillonnage de Gibbs ne fonctionne pas lorsque le nombre de paramètres est variable :

Définissons maintenant la structure de cet algorithme. :

Définition 4.3 :

Etant donné une densité f , une densité g satisfaisant :

$$\int_E g(x, z) dz = f(x), \quad (3.16)$$

Est appelée une complétion de f

Soit f la densité ciblée par l'algorithme. On choisit alors une complétion de f de manière à ce que ses lois conditionnelles soient faciles à simuler. On échantillonnera g plutôt que f . Pour $p > 1$, en notant $y = (x, z)$, les lois conditionnelles de $g(y) = g(y_1, \dots, y_p)$ s'écrivent :

$$g_1(y_1|y_2, \dots, y_p), g_2(y_2|y_1, y_3, \dots, y_p), \dots, g_p(y_p|y_1, \dots, y_{p-1}). \quad (3.17)$$

L'algorithme d'échantillonnage de Gibbs associé à cette décomposition est alors fourni par la transition de $y^{(t)}$ à $y^{(t+1)}$ suivante :

$$\text{Simuler } \mathbf{1} - y_1^{(t+1)} \sim g_1 \left(y_1 | y_2^{(t)}, \dots, y_p^{(t)} \right)$$

$$\mathbf{2} - y_2^{(t+1)} \sim g_2 \left(y_2 | y_1^{(t)}, y_3^{(t)}, \dots, y_p^{(t)} \right)$$

⋮

$$\mathbf{p} - y_p^{(t+1)} \sim g_p \left(y_p | y_1^{(t)}, \dots, y_{p-1}^{(t)} \right)$$

Cette technique de complétion d'une densité f en une densité g telle que f apparaisse comme densité marginale de g est souvent naturelle. On peut noter la similarité de cette approche avec l'algorithme EM de maximisation de la vraisemblance. De même, il faut remarquer que la méthode d'échantillonnage de Gibbs ne requiert aucunement que la complétion de f en g et de x en $y = (x, z)$ soit reliée au problème inférentiel. Ceci signifie que cette façon d'aborder les choses est justifiée techniquement sans qu'il soit nécessaire de la relier au problème réel sous-jacent.

Robert et Casella, 1999 montrent que la plupart des algorithmes de Gibbs vérifient les conditions minimales nécessaire à l'ergodicité de la chaîne. Pour ce qui est des propriétés de convergence de cet algorithme, il faut montrer que la chaîne $(y^{(t)})$ converge vers la distribution g et que $(x^{(t)})$ est une chaîne qui converge vers f . En effet, même si $(y^{(t)})$ est par construction une chaîne de Markov, $(x^{(t)})$ n'en est pas forcément une.

Théorème :

Si la chaîne $(y^{(t)})$ produite par l'algorithme de Gibbs est ergodique, alors la distribution g est stationnaire pour la chaîne $(y^{(t)})$ et f est la distribution limite de la chaîne $(x^{(t)})$.

On pourra trouver la preuve de ce théorème dans *Robert et Casella, 1999*. Afin de compléter le parallèle entre l'échantillonnage de Gibbs et les méthodes de Metropolis-Hastings, on peut citer le théorème suivant. (*Robert et Casella, 1999 p.296.*)

Théorème :

La méthode d'échantillonnage de Gibbs décrite plus haut correspond à la composition de p algorithmes de Metropolis-Hastings de probabilités d'acceptation égales à 1.

Preuve

On peut considérer l'échantillonnage de Gibbs comme la composition de p algorithmes "élémentaires" correspondant chacun à la simulation suivant une des lois conditionnelles de g . Il suffit donc de montrer que chacun de ces algorithmes a une probabilité d'acceptation égale à 1.

Considérons l'étape i . La loi instrumentale s'écrit

$$q_i(y'/y) = \delta_{(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)}(y'_1, \dots, y'_{i-1}, y'_{i+1}, \dots, y'_p) g_i(y'_i | y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p). \quad (3.18)$$

Le rapport définissant la probabilité $\rho(y, y')$ s'écrit alors

$$\begin{aligned} \frac{g(y')q_i(y|y')}{g(y)q_i(y'|y)} &= \frac{g(y')g_i(y_i|y'_1, \dots, y'_{i-1}, y'_{i+1}, \dots, y'_p)}{g(y)g_i(y'_i|y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)} \\ &= \frac{g_i(y'_i|y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)g_i(y_i|y'_1, \dots, y'_{i-1}, y'_{i+1}, \dots, y'_p)}{g_i(y_i|y'_1, \dots, y'_{i-1}, y'_{i+1}, \dots, y'_p)g_i(y'_i|y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)} \\ &= 1 \end{aligned}$$

Considérations pratiques : initialisation, et convergence :

En tant qu'algorithmes MCMC, les diverses méthodes présentées ici nécessitent toutes le choix d'une initialisation $x^{(0)}$. De manière idéale, on devrait choisir $x^{(0)}$ selon la loi invariante f ciblée par l'algorithme. Cela est rarement possible, c'est pourquoi nous utiliserons généralement une valeur au hasard tirée selon la loi a priori. L'influence du point de départ de l'algorithme devient vite négligeable. Toutefois, pour réduire au maximum l'influence de $x^{(0)}$ sur l'estimation de la densité ciblée, il est d'usage de ne pas tenir compte d'un certain nombre m d'itérations, ceci signifiant que les m premières valeurs générées par l'algorithme ne seront tout simplement pas prises en compte. Les auteurs anglophones font alors référence à la période de "burn in" ou "temps de chauffe".

Si la chaîne de Markov générée est capable de mouvements rapides et amples permettant de se déplacer rapidement dans tout l'espace d'état, alors la valeur générée $x^{(m)}$ pourra être considérée comme étant indépendante de $x^{(0)}$ pour m relativement grand.

Les bonnes ou mauvaises propriétés d'estimation de la chaîne déterminent aussi la taille totale N de la chaîne à générer afin d'obtenir une estimation fiable de la densité f ciblée. Plus la chaîne se déplace vite, plus la convergence sera rapide.

Plusieurs méthodes existent permettant de déterminer de manière théorique les ordres de grandeur pour m et T . On pourra par exemple se référer à *Cowles et Carlin, 1996* pour un passage en revue des diverses solutions existantes. On remarquera cependant qu'on ne peut connaître a priori le nombre d'itérations nécessaires à la convergence.

Dans la prochaine section, nous utilisons cette méthode permettant d'estimer les paramètres du modèle présenté dans *le chapitre 2*.

3.4 Estimation des paramètres d'un modèle SV

3.4.1 Estimation

On se place dans la cadre du modèle canonique (2.1). (après transformation sous équations d'espace d'état.) Notons $\theta = (\sigma_\eta, \phi, \beta)$ le vecteur des paramètres et θ_0 la vraie valeur. On observe une trajectoire de longueur N , soit $y = (y_1, \dots, y_T)$, du processus (y_t) .

Réécrivons le modèle :

$$y_t = \beta e^{h_t/2} \varepsilon_t, \quad (2.1.a)$$

$$\begin{aligned} h_{t+1} &= \mu + \phi (h_t - \mu) + \sigma_\eta \eta_t, \\ h_1 &\rightarrow N \left(\mu, \frac{\sigma_\eta^2}{1 - \phi^2} \right). \end{aligned} \quad (2.1.b)$$

Après transformation sous espace d'état

$$y_t^* = h_t + z_t \quad (2.38.a),$$

$$z_t = \log \varepsilon_t^2$$

$$h_{t+1} = \mu + \phi (h_t - \mu) + \sigma_\eta \eta_t \quad (2.38.b),$$

Rappelons que les autres variables du modèle, en particulier la volatilité h_t , ne sont pas observables, pour l'estimation des paramètres de modèle SV on va suivre l'algorithme suivant :

3.4.2 Procédure d'estimation

Nous allons suivre l'algorithme suivant :

1) Initialiser $s^{(0)}, \phi^{(0)}, \sigma_\eta^{2(0)}$ et $\mu^{(0)}$.

2) Pour $m = 1, \dots, M$

1- pour $t = 1, \dots, T$

Echantillonner $h_t^{(m)}$ à partir de $h|y^*, s, \phi, \sigma_\eta^2$.

2- Echantillonner s à partir de $s|y^*, h$. avec $s = (s_1, \dots, s_T)$

3- calculer $\sigma_\eta^{2(m)}$ à partir de $p(\sigma_\eta^2|y, h^{(m)}, \mu^{(m-1)}, \phi^{(m-1)})$.

4- calculer $\phi^{(m)}$, à partir de $p(\phi|y, h^{(m)}, \sigma_\eta^{2(m)}, \mu^{(m-1)})$.

5- calculer $\mu^{(m)}$ à partir de $p(\mu|y, h^{(m)}, \phi^{(m)}, \sigma_\eta^{2(m)})$.

Notons que nous avons utilisé $y^* = \{\log(y_1^2 + c), \dots, \log(y_t^2 + c), \dots, \log(y_T^2 + c)\}$, Les vecteurs y^* et y , contiennent les mêmes informations.

Proposition 4.1 (*Kim et al, 1998*)

Kim et al, 1998 ont supposé que les paramètres du modèle sont a priori indépendamment répartis comme suit :

$$p(\sigma_\eta^2) \sim GI(\alpha_\sigma, \beta_\sigma^2) \quad (3.19.a)$$

$$p(\phi) \propto \left\{ \frac{(1 + \phi)}{2} \right\}^{\phi^{(1)} - 1} \left\{ \frac{(1 - \phi)}{2} \right\}^{\phi^{(2)} - 1}, \quad (3.19.b)$$

$$p(\mu) \sim N(\alpha_\mu, \beta_\mu^2) \quad (3.19.c)$$

Où $GI(.,.)$ dénote une distribution Gamma–Inverse, .et $\phi = 2\phi^* - 1$ tel que $\phi^* \sim B\grave{e}ta(\phi^{(1)}, \phi^{(2)})$,

Les paramètres $\alpha_{(.)}$ et $\beta_{(.)}$, caractérisant les distributions a priori, sont appelés "hyper – parameter", qui sont spécifiés par *Kim et al, 1998*.

■

Proposition 4.2(*Kim et al, 1998*)

Étant donné les distributions a priori, les densités a postérieures des paramètres du modèle sont calculées de la manière suivante :

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2) \quad (3.20.a)$$

$$p(\phi|y, h, \mu, \sigma_\eta^2) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\phi). \quad (3.20.b)$$

$$p(\mu|y, h, \sigma_\eta^2, \phi) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\mu) \quad (3.20.c)$$

■

Puisque l'état de la volatilité h subsume toutes les informations sur les paramètres $(\mu, \phi, \sigma_\eta^2)$, l'information complète de la fonction de vraisemblance $(y|h, \mu, \phi, \sigma_\eta^2)$ est une constante, et par conséquent elle peut être omise.

Successivement nous obtenons :

$$p(h|\mu, \phi, \sigma_\eta^2) \propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) \quad (3.21)$$

Où $p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)$ est donné en (2.1.b). Les densités a posteriori peut être reformulées, après élimination des termes constants. (pour plus de détails, voir ci-dessous).

1) Estimation de σ_η^2 à partir de $p(\sigma_\eta^2|y, h, \mu, \phi)$:

Proposition 4.3 (*Kim et al, 1998*)

La densité à posteriori $p(\sigma_\eta^2|y, h, \mu, \phi)$ est proportionnelle à une densité Gamma–Inverse.

Par conséquent, nous avons :

$$\sigma_\eta^2/y, h, \phi, \mu \sim IG(a, b) \quad (3.22)$$

Si on suppose que $\alpha_\sigma = \frac{\sigma_r}{2}$,
 $\beta_\sigma^2 = \frac{S_\sigma}{2}$

Alors :

$$\begin{aligned} a &= \frac{T}{2} + \alpha_\sigma \\ &= \frac{n + \sigma_r}{2} \\ b &= \beta_\sigma^2 + \frac{(h_1 - \mu)^2 (1 - \phi^2) + \sum_{t=1}^{T-1} ((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2} \\ &= \frac{S_\sigma + (h_1 - \mu)^2 (1 - \phi^2) + \sum_{t=1}^{T-1} ((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2} \end{aligned}$$

On prend $\sigma_r = 5$ et $S_\sigma = 0.01 \times \sigma_r$ ■

Preuve

En utilisant le théorème de Bayes, la distribution à postériori conditionnelle de σ_η^2 est donnée par :

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

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) IG(\sigma_\eta^2|\alpha_\sigma, \beta_\sigma),$$

Tel que la fonction de densité $p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)$ est donnée en (2.1.b).

Après l'élimination de tous les termes constants, on obtient :

$$\begin{aligned} & p(\sigma_\eta^2|y, h, \mu, \phi) \\ & \propto \exp \left[-\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1} \{h_{t+1} - \mu - \phi(h_t - \mu)\}^2}{2\sigma_\eta^2} \right] \\ & \times \left(\frac{1}{\sigma_\eta^2} \right)^{\frac{T}{2}} \frac{(\beta_\sigma)^{\alpha_\sigma} \exp\left(\frac{-\beta_\sigma}{\sigma_\eta^2}\right)}{\Gamma(\alpha_\sigma) (\sigma_\eta^2)^{\alpha_\sigma+1}} \\ & \propto \exp \left[-\frac{\beta_\sigma + \frac{1}{2}(h_1 - \mu)^2(1 - \phi^2) + \frac{1}{2} \sum_{t=1}^{T-1} \{h_{t+1} - \mu - \phi(h_t - \mu)\}^2}{\sigma_\eta^2} \right] \\ & \times \left(\frac{1}{\sigma_\eta^2} \right)^{(\alpha_\sigma + \frac{T}{2})+1} \end{aligned}$$

2) Estimation de ϕ à partir de $p(\phi|y, h, \mu, \sigma_\eta^2)$:
Proposition 4.2 (Kim et al, 1998)

Pour ϕ , nous avons brièvement illustré une procédure d'échantillonnage qui est également utilisée par Kim et al.(1998). L'échantillonnage de la densité conditionnelle est fait par l'utilisation de l'algorithme Metropolis-Hasting.

Soit $\phi = 2\phi^* - 1$ tel que ϕ^* est réparti comme une distribution Bêta des paramètres $(\phi^{(1)}, \phi^{(2)})$, (par la suite on prend $\phi^{(1)} = 20$ et $\phi^{(2)} = 1.5$).

Par conséquent, la densité a priori de ϕ est proportionnelle à :

$$p(\phi) \propto \left\{ \frac{(1+\phi)}{2} \right\}^{\phi^{(1)}-1} \left\{ \frac{(1-\phi)}{2} \right\}^{\phi^{(2)}-1}, \quad \phi^{(1)}, \phi^{(2)} > \frac{1}{2} \quad (3.23)$$

Ce qui donne

$$p(\phi|y, h, \mu, \sigma_\eta^2) \propto p(h|\mu, \phi, \sigma_\eta^2)p(\phi). \quad (3.24)$$

De telle sorte que :

$$p(h|\mu, \phi, \sigma_\eta^2) \propto \exp \left\{ \frac{-(h_1 - \mu)^2 (1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log (1 - \phi^2) \right. \\ \left. - \frac{\sum_{t=1}^{T-1} \{(h_t - \mu) - \phi(h_1 - \mu)\}^2}{2\sigma_\eta^2} \right\} \quad (3.25)$$

■

On note :

$$g(\phi) = \log p(\phi) - \frac{(h_1 - \mu)^2 (1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log (1 - \phi^2) \quad (3.26)$$

La méthode de *Chib et Greenberg 1994*, qui est basée sur l'algorithme de *Metropolis – Hastings*, est appliquée de sorte à échantillonner cette distribution. Plus précisément, étant donné la valeur actuelle $\phi^{(i-1)}$ à la $(i-1)$ ème itération, nous échantillons la valeur ϕ^* à partir de la distribution normale proposée $N(\hat{\phi}, V_\phi)$, où

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

Ensuite, accepter cette proposition comme une valeur $\phi^{(i)}$ avec la probabilité $\exp \left\{ g(\phi^*) - g(\phi^{(i-1)}) \right\}$

Si la valeur proposée est rejetée, alors $\phi^{(i)}$ est égal à $\phi^{(i-1)}$.

3) Estimation de μ à partir de $p(\mu|y, h, \sigma_\eta^2, \phi)$:
Proposition 4.3 (*Kim et al, 1998*)

Le paramètre μ est estimé à partir de la densité à posteriori donnée par la formule suivante

$$\mu|y, h, \phi, \sigma_\eta^2 \sim N(\hat{\mu}, \sigma_\mu^2) \quad (3.27)$$

Telque :

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

Et

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

On considère $\beta = \exp(\mu/2)$

Pour des raisons techniques, *Kim et al, 1998* ont pris : $\alpha_\mu = 0$, et $\beta_\mu^2 = 10$ (i.e. la densité a priori $\mu \sim N(0, 10)$)

Preuve

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

$$\begin{aligned} &\propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{n-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) N(\alpha_\mu, \beta_\mu), \\ &\propto \exp \left(-\frac{1}{2} \left[\underbrace{\mu^2 \left\{ \frac{1 - \phi^2 + (T - 1)(1 - \phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2} \right\}}_A \right. \right. \\ &\quad \left. \left. - 2\mu \underbrace{\left\{ \frac{h_1(1 - \phi^2) + (1 - \phi) \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2} \right\}}_B \right] \right) \\ &\propto N \left(\frac{B}{A}, \frac{1}{A} \right) \end{aligned}$$

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

Remarque

1. Pour l'estimation de h_t On applique la méthode du filtre de kalman détaillée dans le chapitre 2
2. En utilisant la fonction de probabilité suivante pour l'échantillonnage de s_t

$$Pr(s_t = i / y_t^*, h_t) \propto q_i f_N(y_t^* | h_t + m_i - 1.2704, v_i^2), i \leq K$$

Chapitre 4

Application numérique

4.1 Simulation et Estimation d'un modèle SV

Les résultats qui seront présentés dans cette section sont basés sur des données simulées. En effet, on a simulé la variable h_t , à partir de l'équation suivante :

$$\begin{aligned} h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \\ h_1 &\rightarrow N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right) \end{aligned}$$

En deuxième lieu, on a simulé y_t , à partir de :

$$y_t = \beta e^{h_t/2} \varepsilon_t,$$

Enfin, on simule la variable latente y^* comme suit :

$$y_t^* = h_t + z_t, \quad \text{avec } z_t = \log(\varepsilon_t^2) \quad (\text{un mélange gaussien})$$

Où la vraie valeur du paramètre $(\phi, \sigma, \mu, \beta) = (0.96, 0.15, -0.88, 0.6440)$ pour le modèle $M1$ et $(\phi, \sigma, \mu, \beta) = (0.7, 0.15, -0.88, 0.6440)$ pour le deuxième modèle $M2$

Nous nous servirons ensuite de ces données artificielles (estimées) pour estimer les pa-

paramètres du modèle supposés inconnus en utilisant l'algorithme MCMC présenté dans le chapitre 3, que nous avons implémenté sous l'environnement Matlab 7.0. Pour chaque cas, le processus est simulé avec des échantillons de tailles 100, 500, 700 et 1000, pour un nombre de réplique égale à 200,

Nous avons généré des chaînes de taille $M = 15000$ et on élimine les $m = 5000$ premières itérations. On pourra noter que l'algorithme converge plus rapidement et ne nécessite que 10000 à 15000 itérations. Les résultats sont donnés dans les tableaux (1) et (2) tel que :

La colonne (1) contient les vraies valeurs des paramètres ;

La colonne (2) contient les diverses estimations ; autrement dit les moyennes a postériori

La colonne (3) contient les écarts types.

Nous tenons à préciser aussi que nous voulons étudier l'influence ϕ lorsqu'il est proche de la zone de non stationnarité sur l'estimation des paramètres.

Modèle M1 :**N = 100**

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.96	0.9012	0.09531
σy	0.15	0.14395	0.03531
μy	-0.88	-0.81398	0.331288
βy	0.6440	0.668389	0.11638

N = 500

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.96	0.95012	0.095
σy	0.15	0.1457	0.0991
μy	-0.88	-0.904009	0.129
βy	0.6440	0.6364	0.1013

N = 700

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.96	0.95598	0.088
σy	0.15	0.152070	0.098
μy	-0.88	-0.8582	0.051
βy	0.6440	0.6511	0.073

N = 1000

	(1)	(2)	(3)
	Vrai valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.96	0.96033	0.005
σy	0.15	0.15009	0.017
μy	-0.88	-0.8785	0.014
βy	0.6440	0.6445	0.032

Modèle M2 :

Table (2)
N = 100

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.70	0.8549	0.25185
σy	0.15	0.14784	0.29931
μy	-0.88	-0.7094	0.3588
βy	0.6440	0.70197	0.18972

N = 500

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.70	0.7955	0.1985
σy	0.15	0.1477	0.19597
μy	-0.88	-0.8266	0.1792
βy	0.6440	0.6616	0.11007

N = 700

	(1)	(2)	(3)
	Vrai Valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.70	0.7517	0.0993
σy	0.15	0.1472	0.0125
μy	-0.88	-0.8167	0.12951
βy	0.6440	0.6647	0.09713

N = 1000

	(1)	(2)	(3)
	Vrai valeur	Valeur Estimée par MCMC	Écart type std
ϕy	0.70	0.7446	0.015
σy	0.15	0.15009	0.0017
μy	-0.88	-0.8748	0.0195
βy	0.6440	0.6457	0.0062

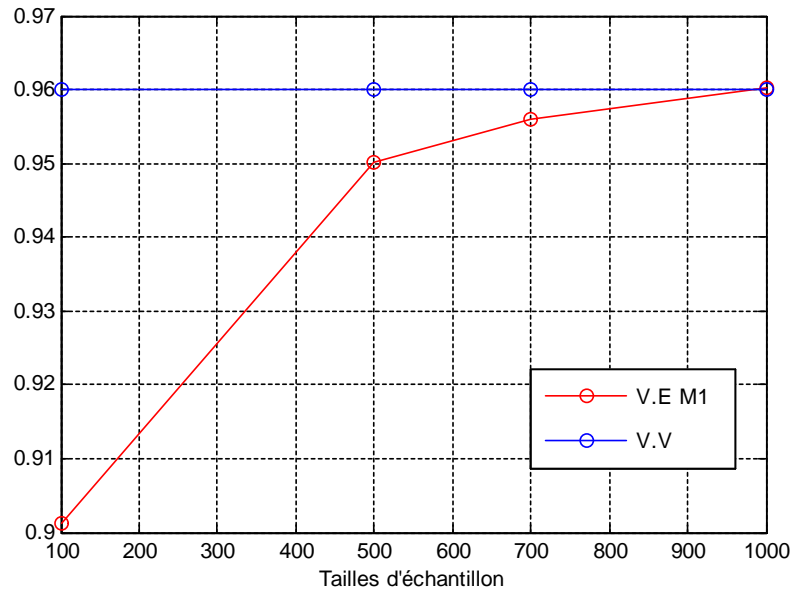


Figure 1 : Valeurs estimées de ϕ pour $\phi = 0.96$

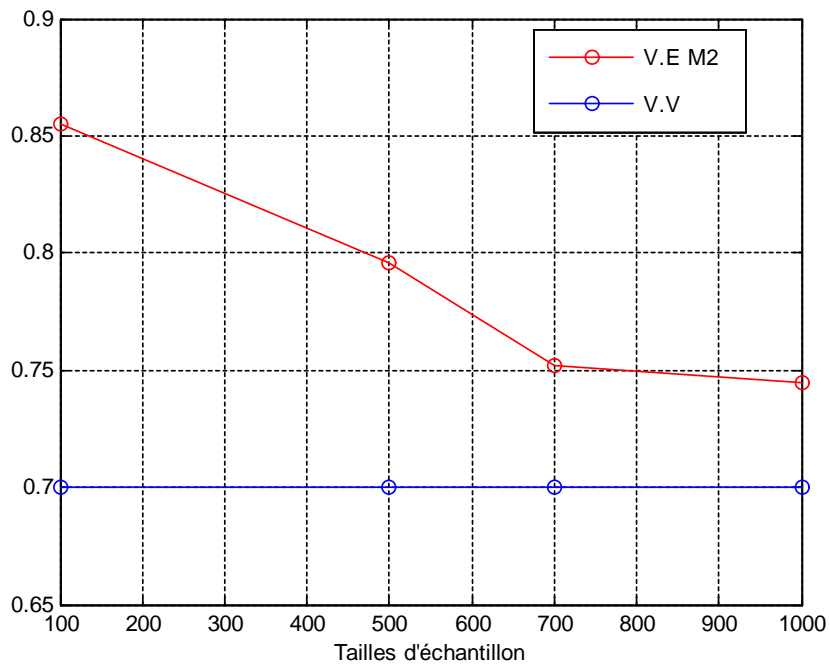


Figure 2 : Valeurs estimées de ϕ pour $\phi = 0.70$

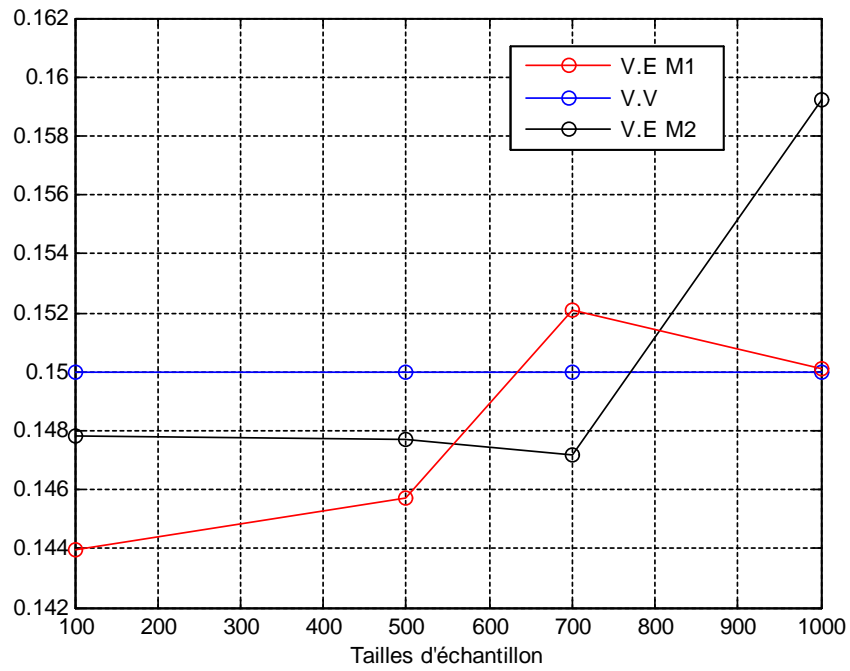


Figure 3 : Valeurs estimées de σ pour $\phi = 0.96$ et $\phi = 0.70$

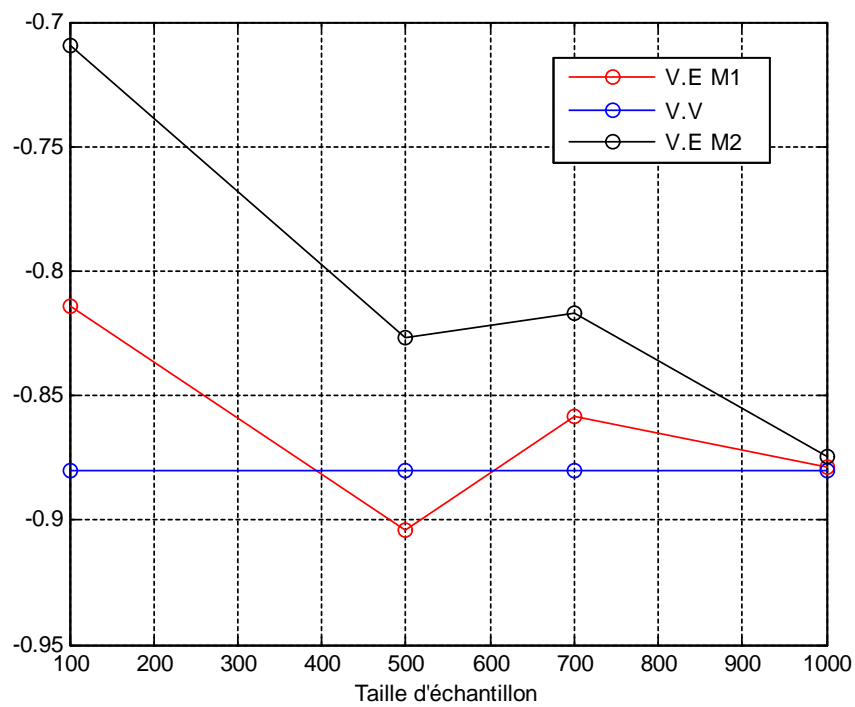


Figure 4 : Valeurs estimées de μ pour $\phi = 0.96$ et $\phi = 0.70$

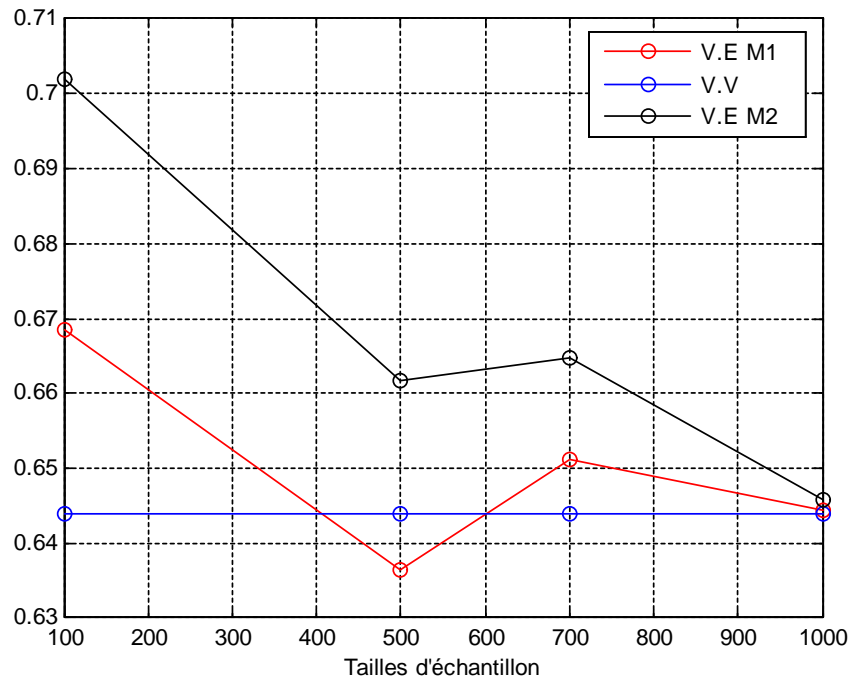


Figure 5 : Valeurs estimées de $\beta = \exp(\mu/2)$ pour $\phi = 0.96$ et $\phi = 0.70$

Remarque

Les notations dans la légende désigne :

- V.E : désigne la valeur estimée du paramètre qui n'est autre que la moyenne a postériori pour le modèle M1 et le modèle M2.

- V.V : désigne la vraie valeur du paramètre que nous avons estimer

Commentaire

Il peut être remarqué que pour chaque instance, les vraies valeurs sont bien estimées particulièrement le paramètre ϕ et les σ ; Cependant μ ainsi que β sont relativement moins précises. Comme on peut remarqué aussi que la qualité des estimateurs, devient relativement plus précise lorsque la taille de l'échantillon augmente ce qui est bien illustré par la suite et qui est confirmé par les résultats donnés dans la **Table (3)** et **Table (4)** correspondant à la deuxième expérience dans laquelle on applique notre algorithme sur des données réelle

Il faut noter aussi que les temps de calcul ont été élevé.

Nous remarquons que les graphes (voir **Figure 1** , **Figure 2** et **Figure 3**) reflétant

l'estimation de ϕ et σ n'ont pas la même allure. mais ceci n'est pas probablement dû qu'aux fluctuations des échantillons de base, car dans tout les cas , les valeur estimées ne semble pas très éloignées des vraies valeurs des paramètres respectifs surtout si la taille de l'échantillon est grande.

Les graphes **Figure 4** et **Figure 5** représentant l'estimation de μ et β ont a peu près la même allure, il se caractérise par une certaine instabilité lorsque T est relativement faible

On peut noter que la convergence vers la vraie valeur du parametre est plus rapide lorsque ϕ est plus proche de la zone de stationnarité.

4.2 Application sur une série de données réelle :

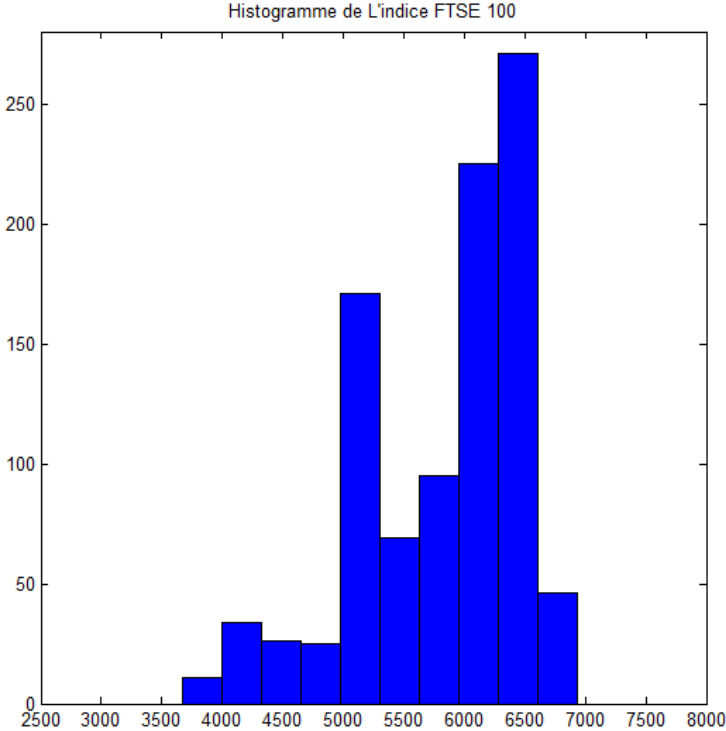
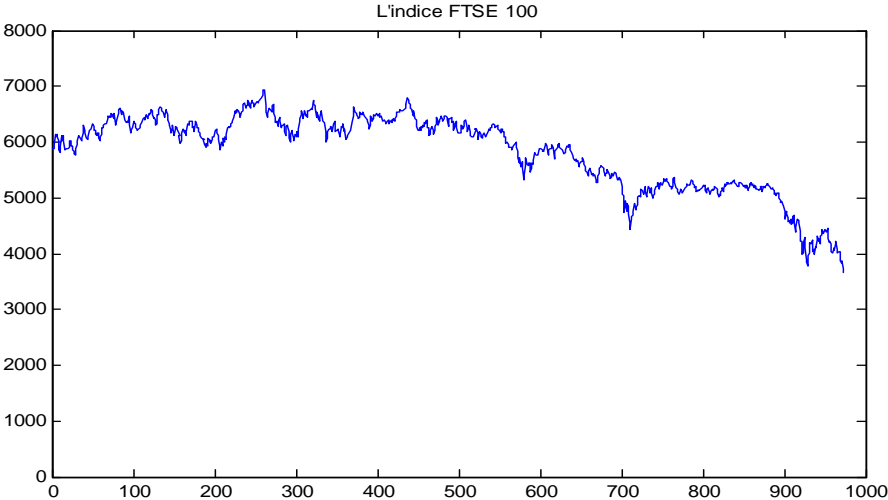
Cette section comporte la description des données (*FTSE* 100) ainsi que les données (*Nikkei* 225) sur lesquelles nous allons travailler, la motivation d'utiliser les modèles à Volatilité Stochastique et les résultats de l'estimation et des prévisions. (où les données sont journalières, 974 observations pour chaque série qui vont du 01 Janvier 1999 au 24 Septembre 2002).

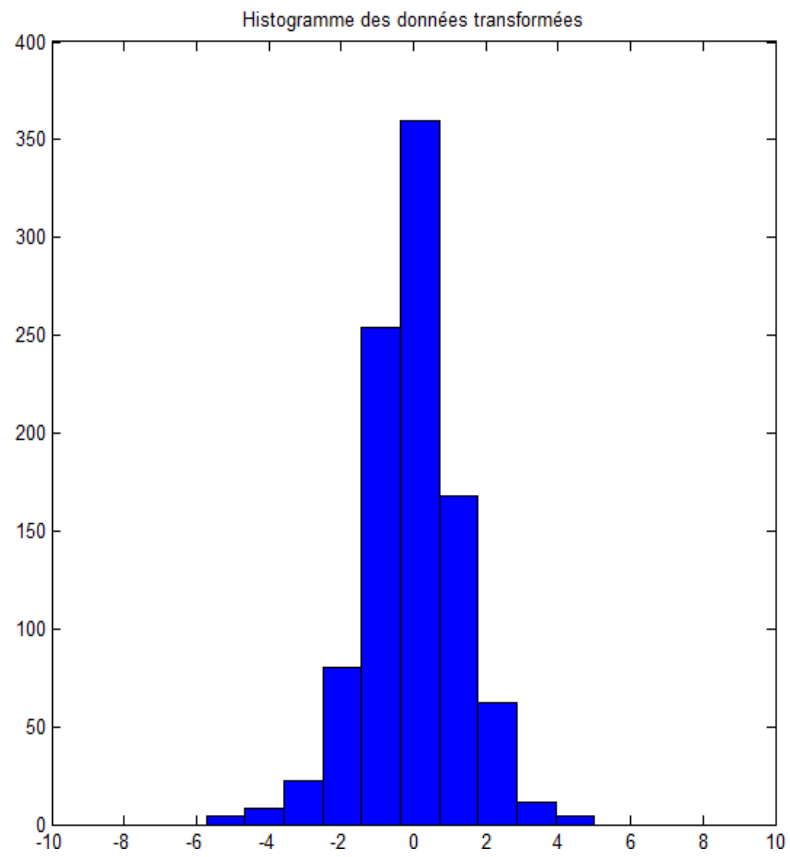
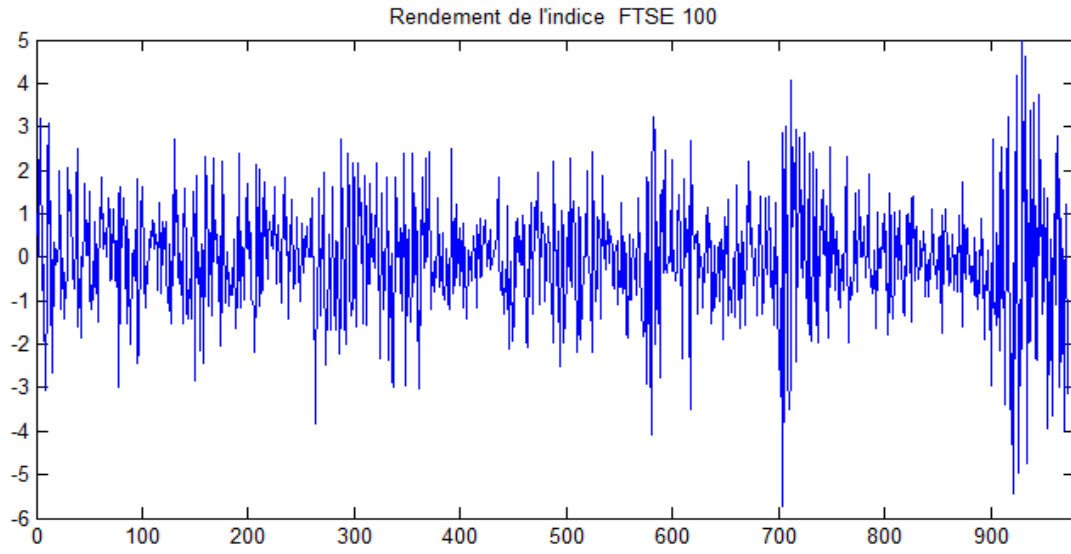
Par la suite tous les résultats seront comparés avec ceux obtenus par *Krichene* (2003) dans son article "Modelling stochastic volatility with application to Stock Returns"

4.2.1 Description des données *FTSE* 100

L'indice *FTSE* 100 que l'on appelle également *FTSE*, ou, de façon familière, le « *footsie* » est un indice boursier des cent entreprises britanniques les mieux capitalisées cotées à la bourse de Londres. Les quatre initiales signifient *Financial Times Stock Exchange*. L'indice a été lancé le 3 janvier 1984 sur une base de 1000 points ; son plafond historique (en 2010) a été atteint le 30 décembre 1999 avec 6950,6 points.

Le *FTSE* 100 est l'indice boursier le plus largement utilisé de tous les indices proposés par le *FTSE Group*, et est fréquemment usité (notamment dans les informations financières) comme une mesure de la bonne santé de l'économie.





Estimation des paramètres

On a généré des chaînes de taille $M = 25000$ et élimine les $m = 5000$ premières itérations.

Les résultats ont été comparés à ceux obtenus par Krichene (2003).

Ligne (1) : nos estimations autrement dit les moyenne a postériori

Ligne (2) : estimations obtenus par *Krichene*

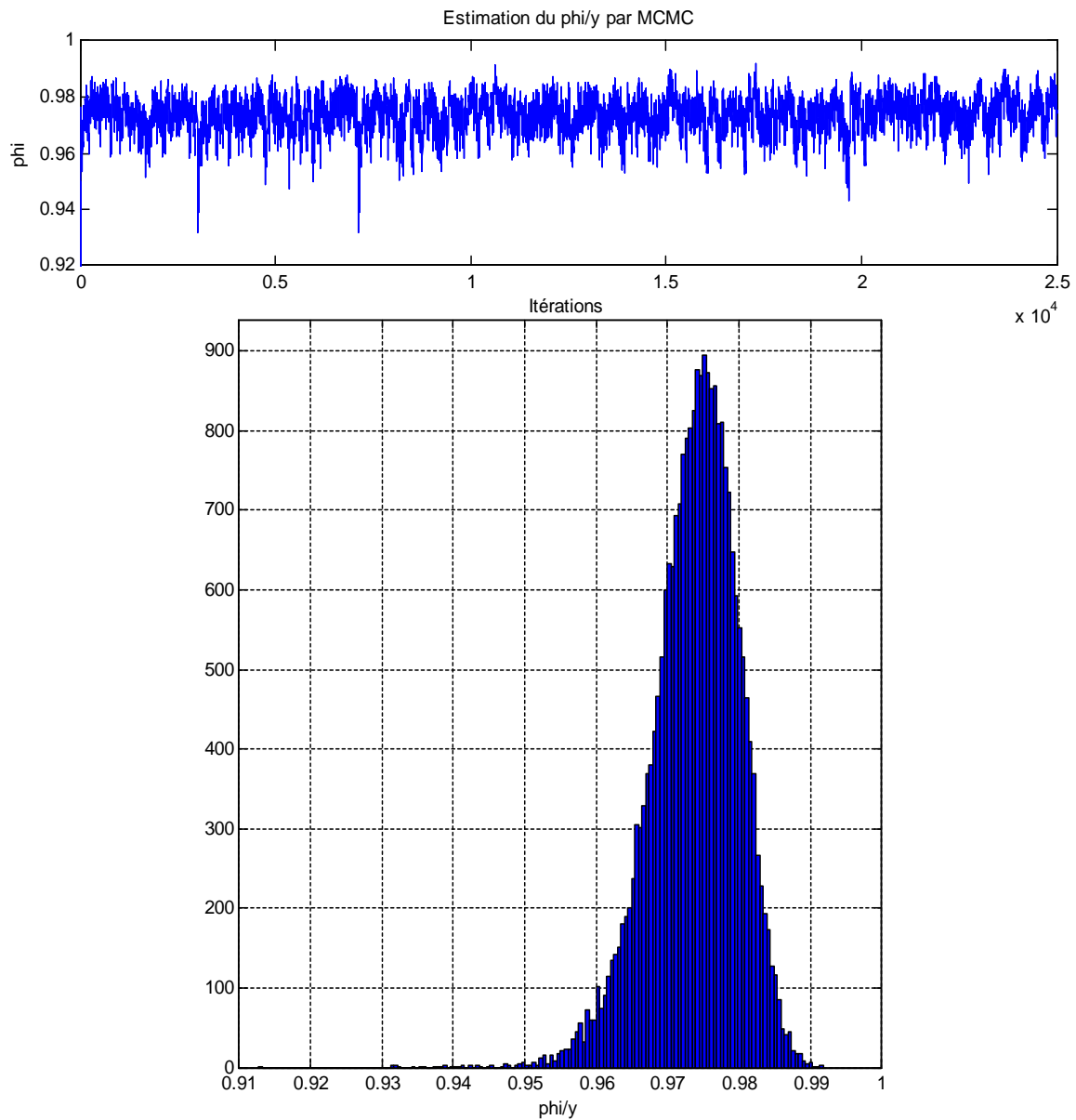
Ils sont résumés dans les tableaux suivants :

Table (3)				
		valeur estimée par MCMC	Écart type std	mse
ϕy	{ (1)	0.9739	0.0060	3.2858×10^{-7}
	{ (2)	0.974		
σy	{ (1)	0.1576	0.01882	6.9996×10^{-8}
	{ (2)	0.157		
μy	{ (1)	0.2999	0.05482	0.0081
	{ (2)	0.334		
βy	{ (1)	1.1622	0.031	6.9247×10^{-4}
	{ (2)	1.182		

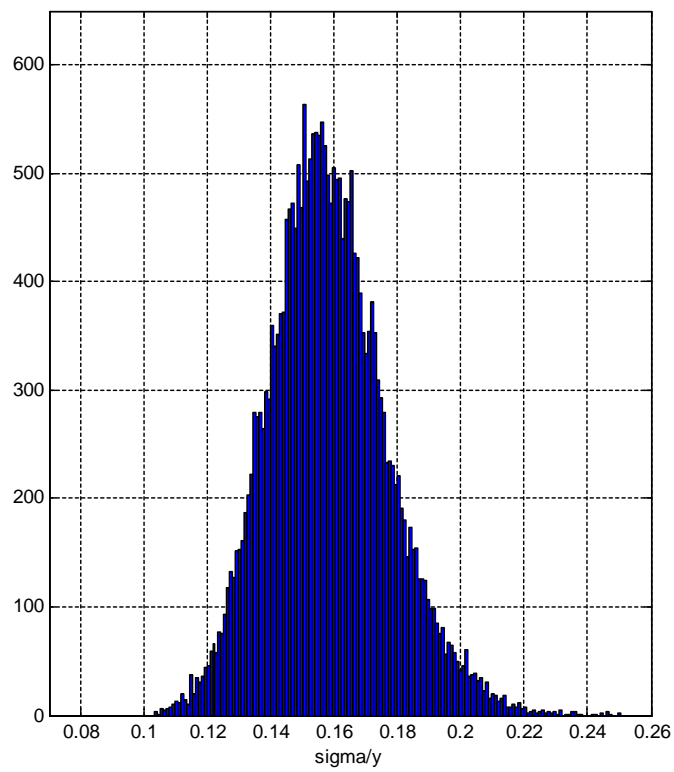
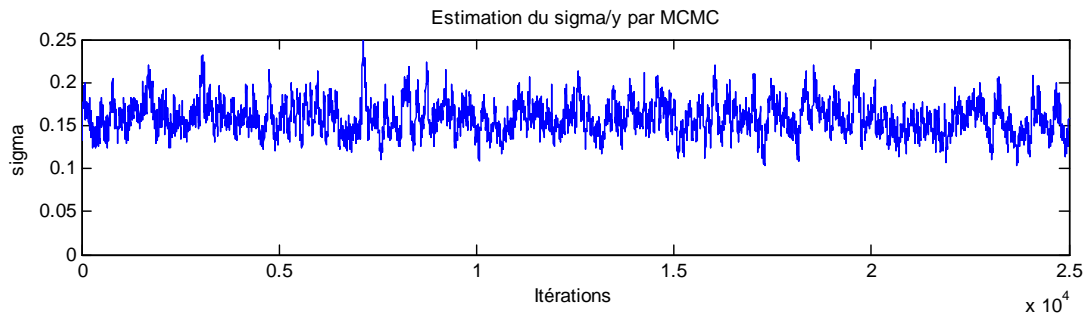
Elapsed time is 10415.902000 seconds.

Graphiques :

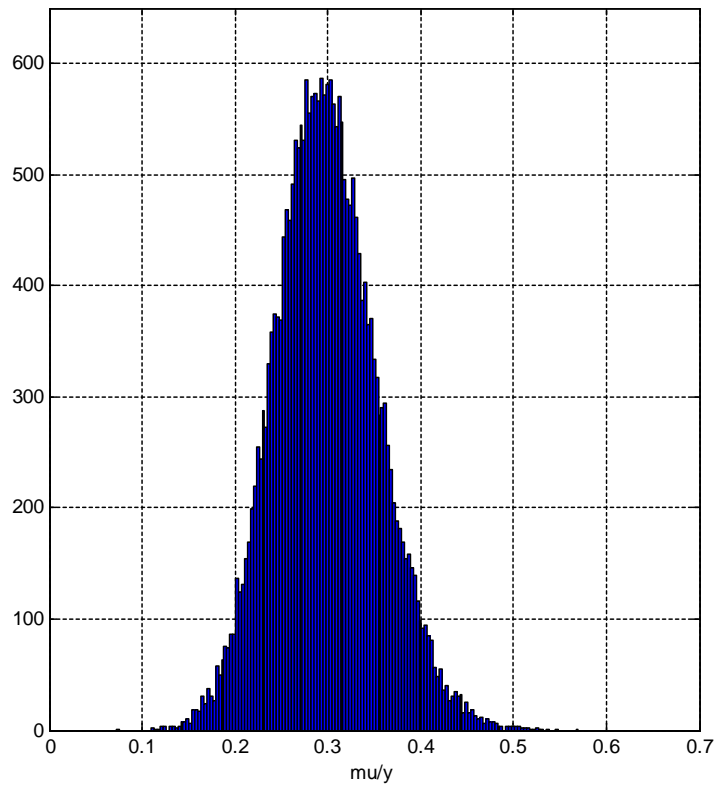
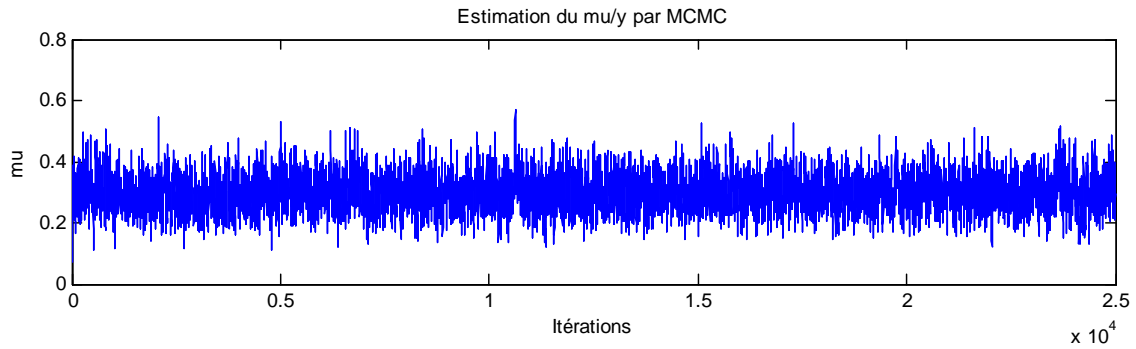
Les figures ci-après représentent les traces des paramètres pour $M = 25.000$ ainsi que leurs distributions à posteriori.

1) Estimation de ϕ à partir de $p(\phi|y, h, \mu, \sigma_\eta^2)$ 

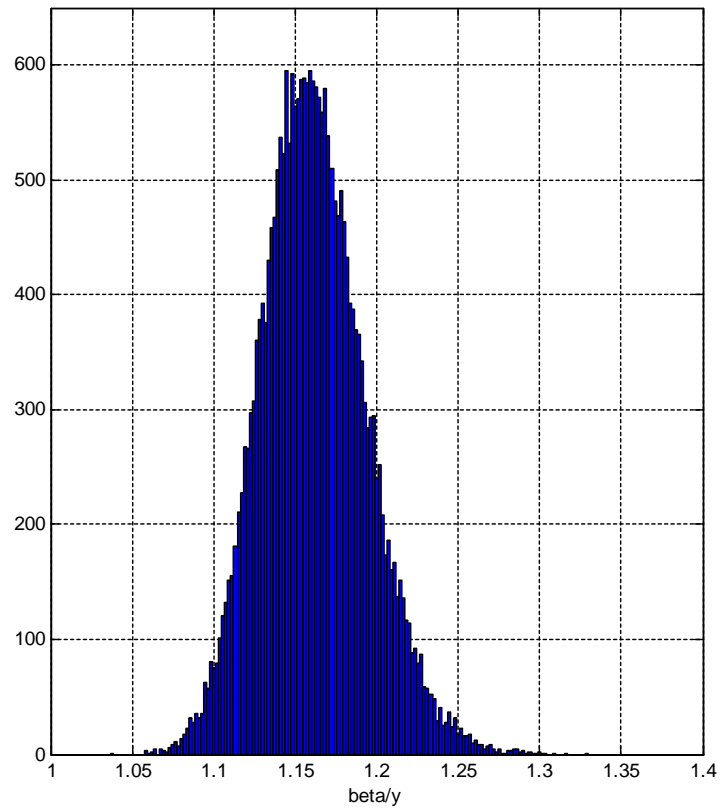
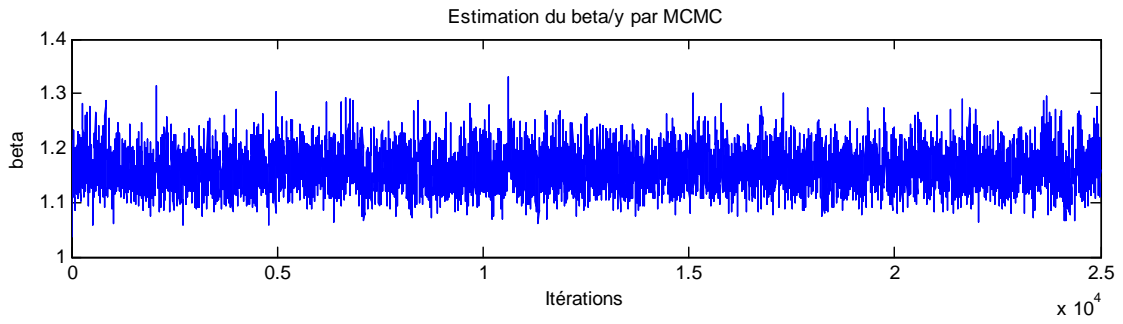
2) Estimation de σ_η^2 à partir de $p(\sigma_\eta^2|y, h, \mu, \phi)$



3) Estimation de μ à partir de $p(\mu|y, h, \sigma_\eta^2, \phi)$



4) Estimation de $\beta = \exp(\mu/2)$

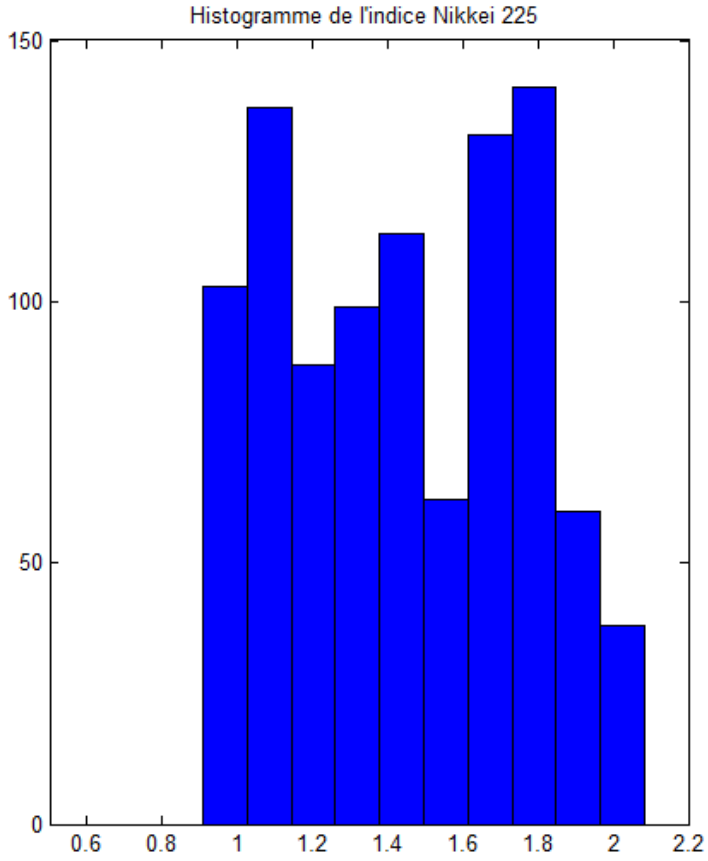
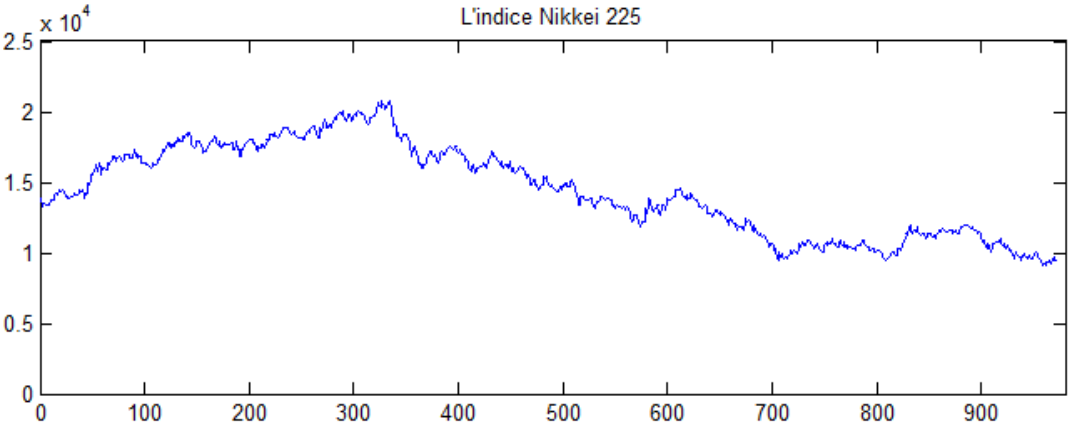


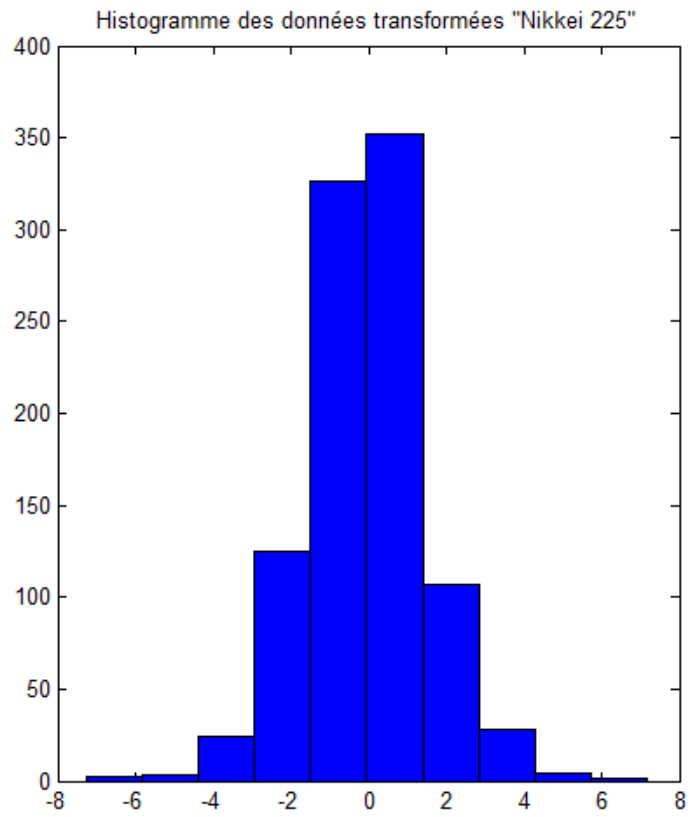
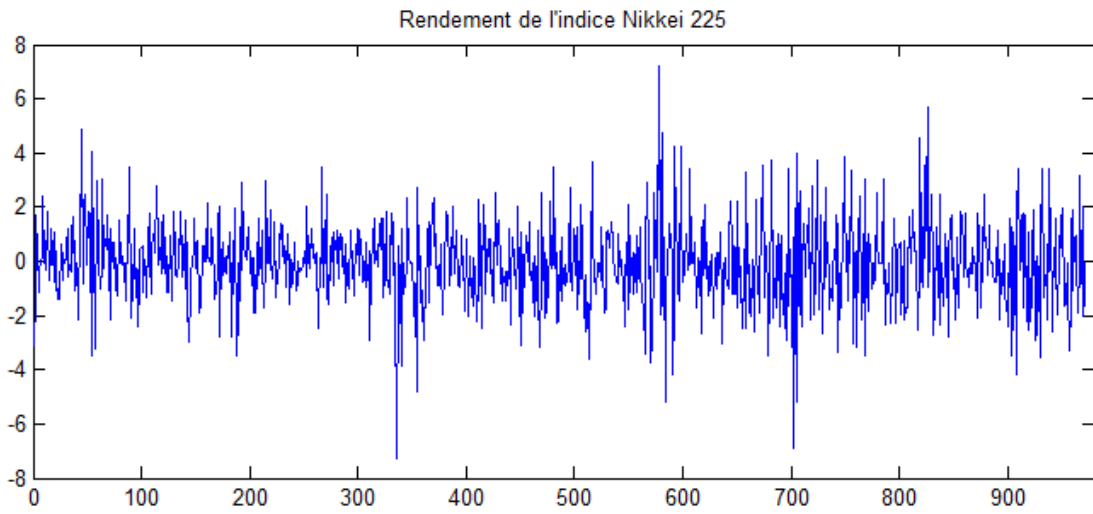
4.2.2 Description des données *Nikkei 225*

Le *Nikkei 225* est le principal indice boursier de la bourse de *Tokyo*. Le terme *Nikkei* est l'abréviation de « *Nihon Keizai Shinbun* », le nom du quotidien économique qui publie cet indice. Il est composé de 225 sociétés.

Le *Nikkei* a été créé le 16 mai 1949 (année de la base 100). Le *Nikkei 225* se calcule par une moyenne arithmétique des valeurs qui le compose, sans pondération par la capitalisation boursière des titres qui le composent.

Cet indice se trouvait à près de 40000 points à la fin 1989 et à 15000 en 1992, suite à l'explosion de la bulle financière. Entre 2000 et 2003, il a accentué sa chute pour atteindre un plus bas de vingt ans. À la faveur du retour en grâce de l'économie nipponne après une stagnation longue d'une décennie, l'indice a repris du terrain, clôturant fin 2005 à plus de 16000 points, à la suite des seconds semestres 2004 et 2005 qui ont été le théâtre de rallies boursiers particulièrement marqués





Estimation du paramètres

Les résultats ont été comparés à ceux obtenus par Krichene (2003).

Ligne (1) : nos estimations autrement dit les moyennes a postériori

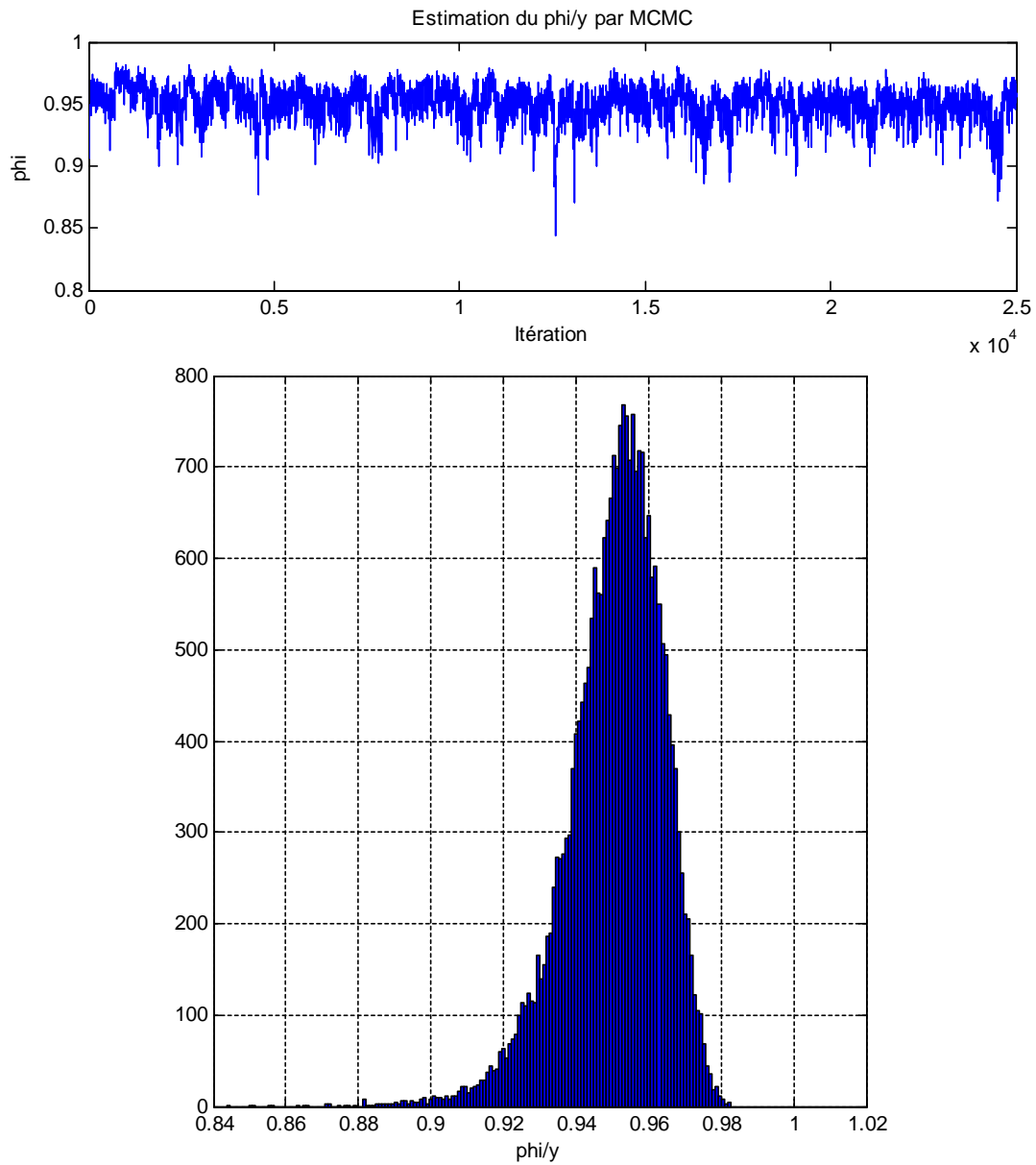
Ligne (2) : estimations obtenus par *Krichene*

Ils sont résumés dans les tableaux suivants [1].

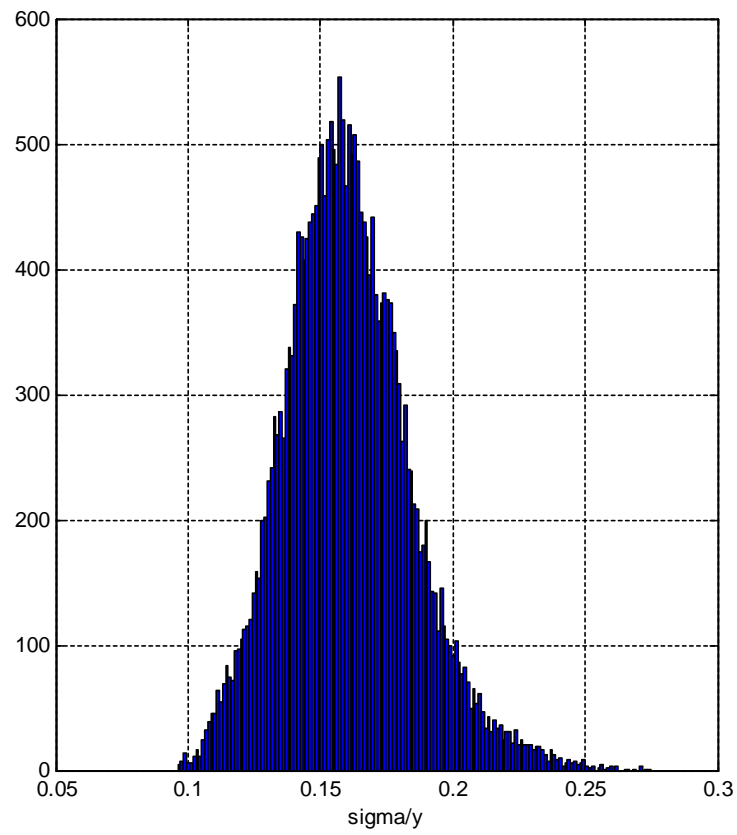
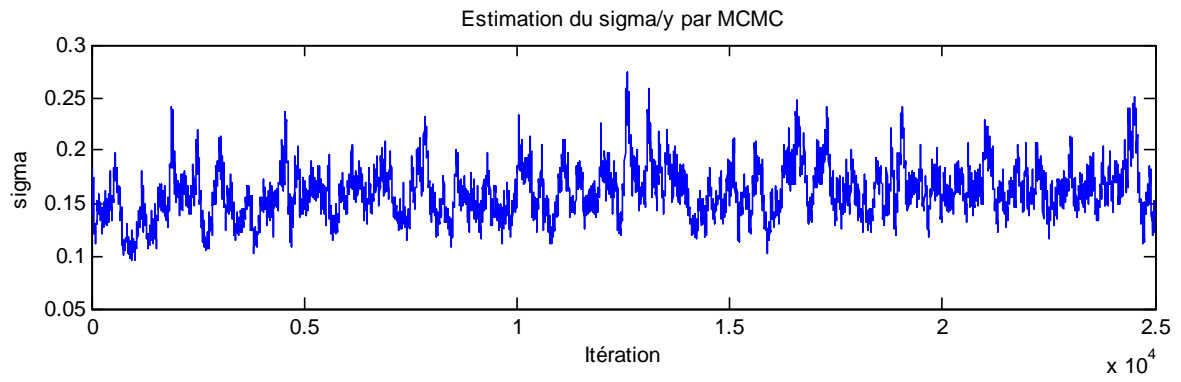
Table (4)				
		valeur estimée par MCMC	Écart type std	Erreur quadratique moyenne mse
ϕy	{ (1)	0.9463	0.01503	1.87231×10^{-10}
	{ (2)	0.944		
σy	{ (1)	0.1706	0.02784	7.29322×10^{-7}
	{ (2)	0.178		
μy	{ (1)	0.7027	0.04763	0.24384
	{ (2)	0.6543		
βy	{ (1)	1.42139	0.0338	0.03153
	{ (2)	1.387		

Elapsed time is 10577.585000 seconds.

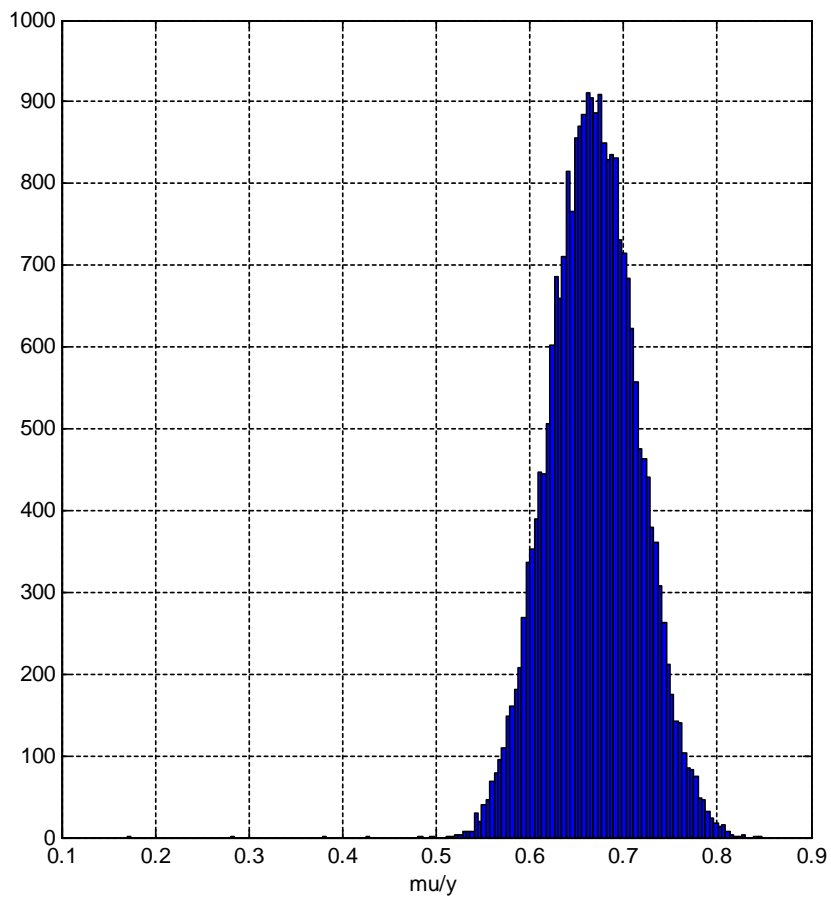
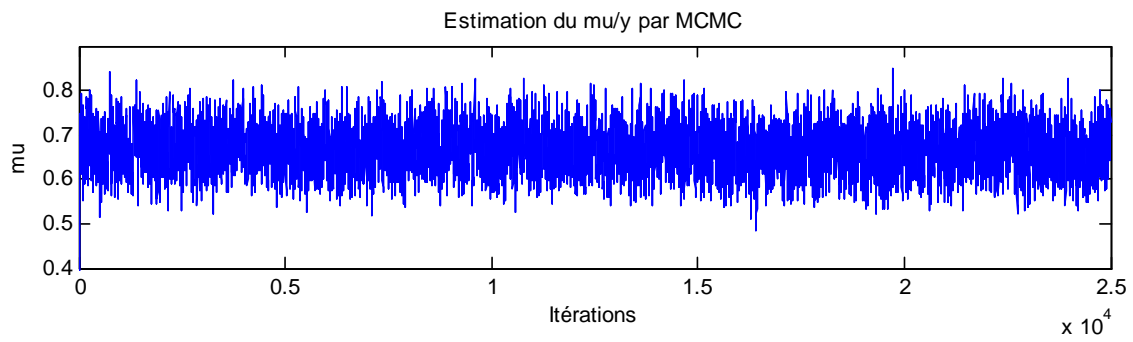
1) Estimation de ϕ à partir de $p(\phi|y, h, \sigma_\eta^2, \mu)$

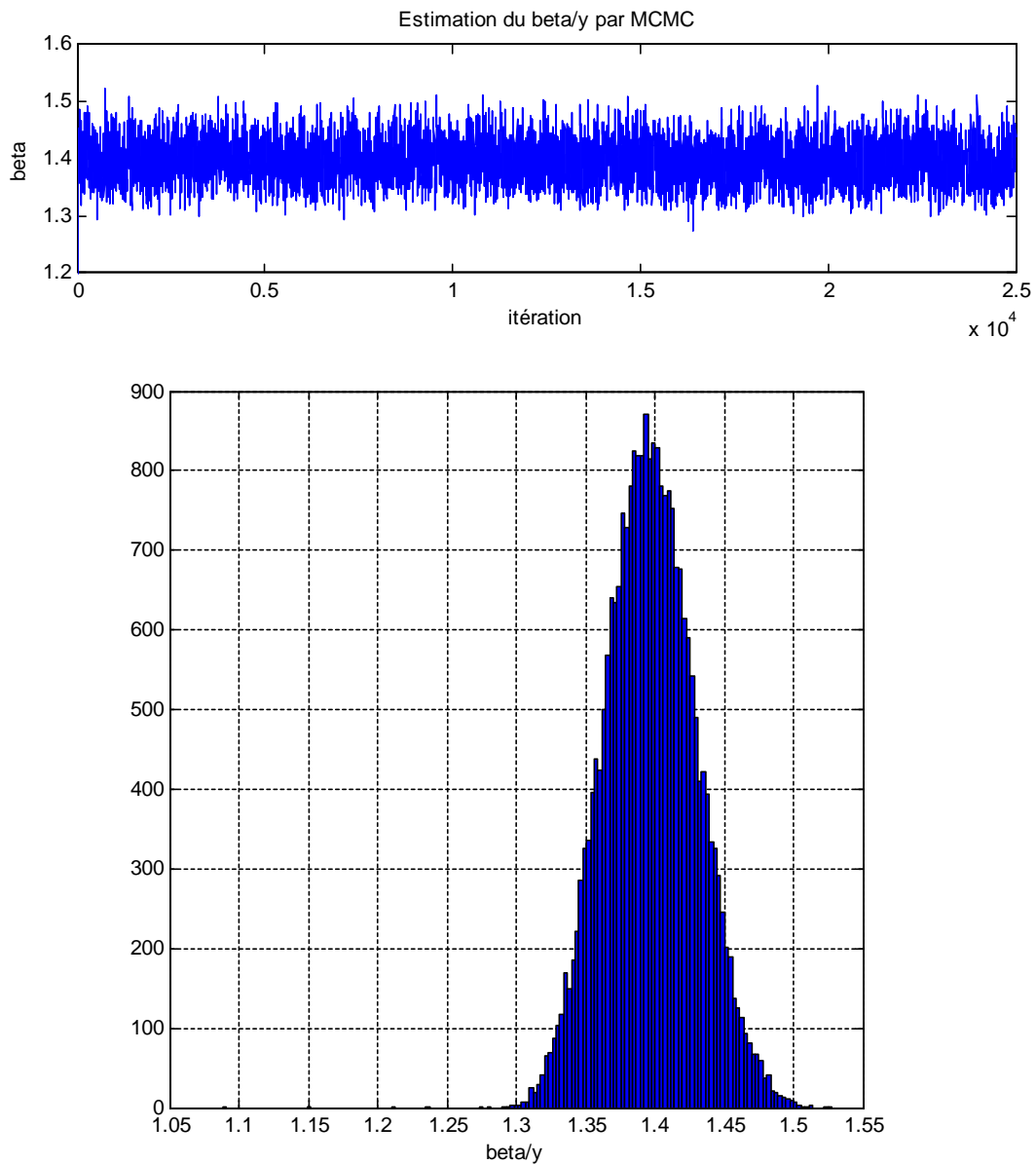


2) Estimation de σ_η^2 à partir de $p(\sigma_\eta^2|y, h, \mu, \phi)$



3) Estimation de μ à partir de $p(\mu|y, h, \sigma_\eta^2, \phi)$



4) Estimation de $\beta = \exp(\mu/2)$ **Commentaire**

Krichene (2003) a appliqué aussi la méthode MCMC dans ce modèle avec une vraie valeur du paramètre égale à celle estimée dans la littérature pour des séries temporelles de données économiques. On remarque que nous avons obtenu les même valeur des estimations pour la séries des données *FTSE 100* et les données *Nikkie 22*; elle converge vers les vraies valeurs que nous avons utilisé comme valeurs initiales

4.2.3 Ajustement de la série *FTSE* 100 par le modèle **SV**

Nous avons ajusté la série à un modèle SV en utilisant notre programme informatique écrit en langage MATLAB. Le modèle estimé est comme suit :

$$y_t = 1.1622 \times e^{h_t/2} \varepsilon_t,$$

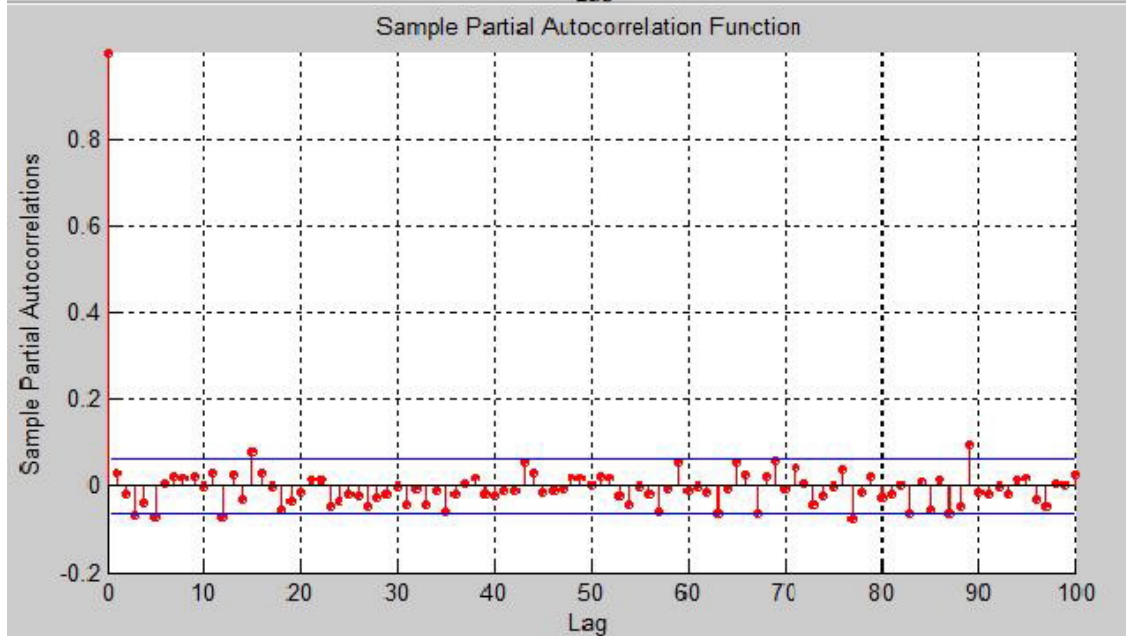
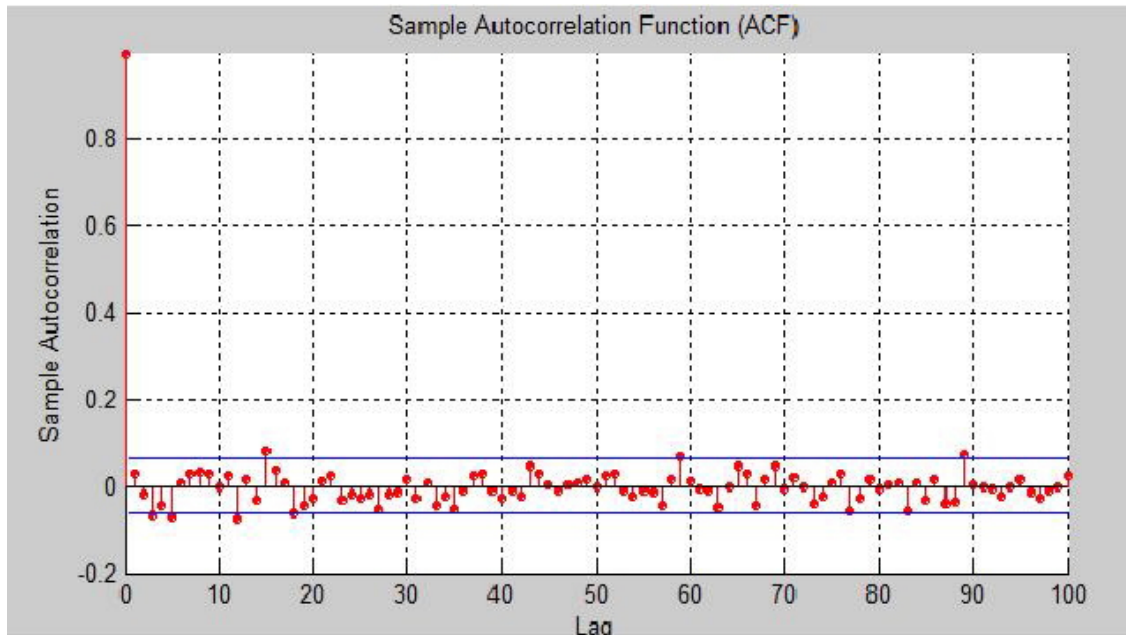
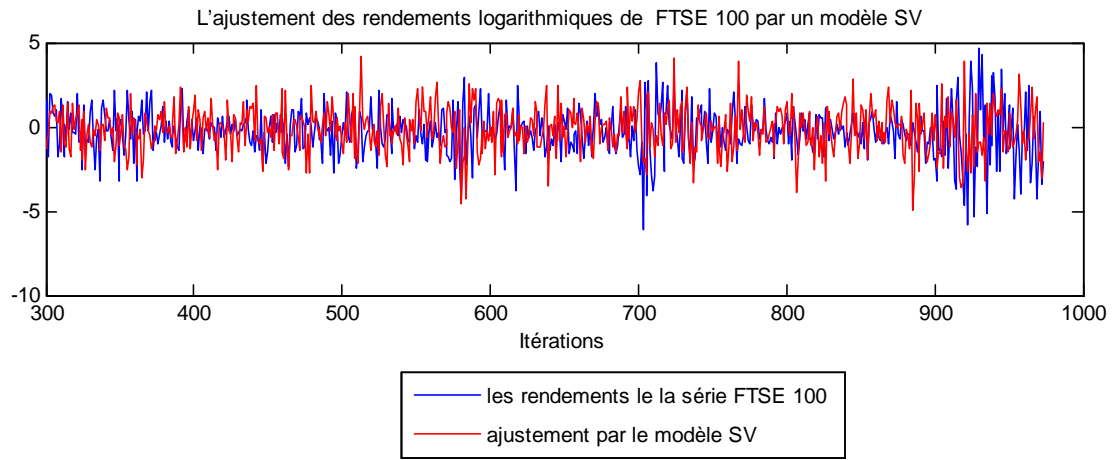
$$h_{t+1} = 0.2999 + 0.9739(h_t - 0.2999) + 0.1576\eta_t$$

$$h_1 \rightarrow N\left(0.2999, \frac{0.1576^2}{1 - 0.9739^2}\right)$$

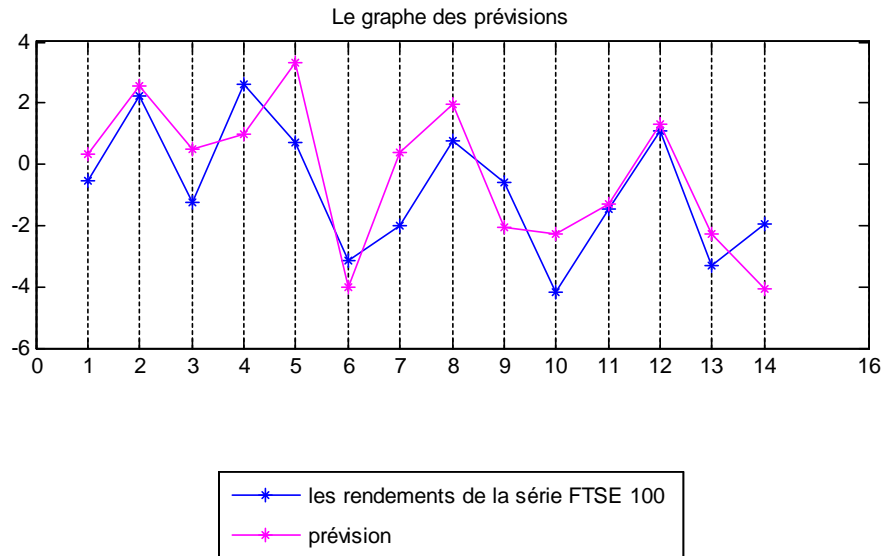
$$\varepsilon_t \rightarrow N(0, 1)$$

$$\eta_t \rightarrow N(0.1)$$

Les figures sont une représentation conjointe de la série des rendements logarithmiques et de la série ajustée et les résidus correspondant.



La figure précédente montre le graphe des prévisions obtenues à partir du modèle SV en plus de la série transformée.



4.2.4 Ajustement de la série *Nikkie 225* par le modèle SV

Nous avons ajusté la série à un modèle SV en utilisant notre programme informatique écrit en langage MATLAB. Le modèle estimé est comme suit :

$$y_t = 1.421398 \times e^{h_t/2} \varepsilon_t,$$

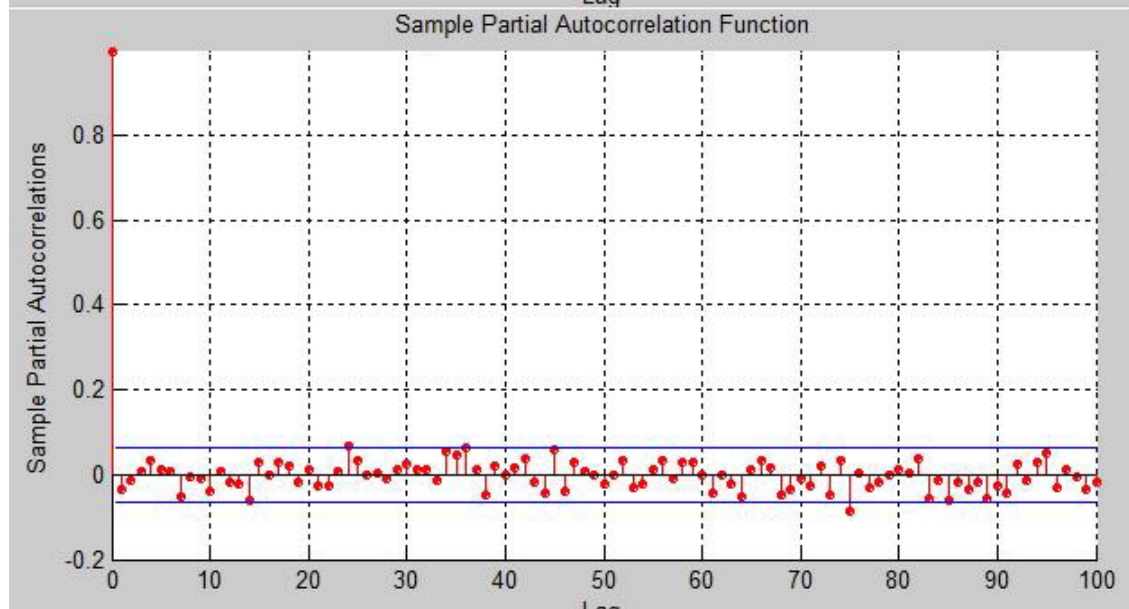
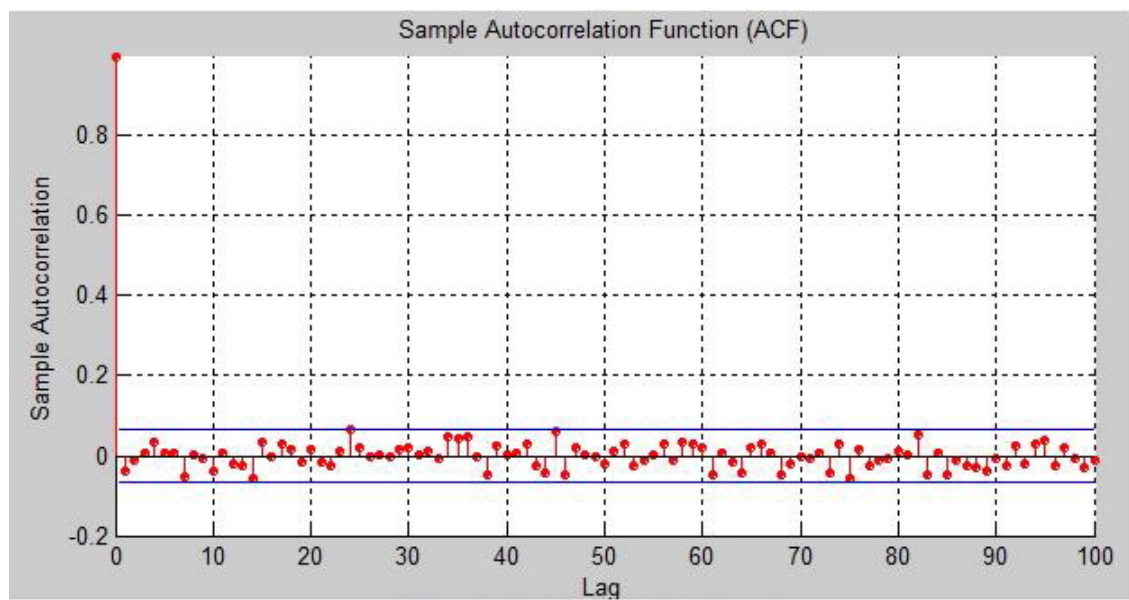
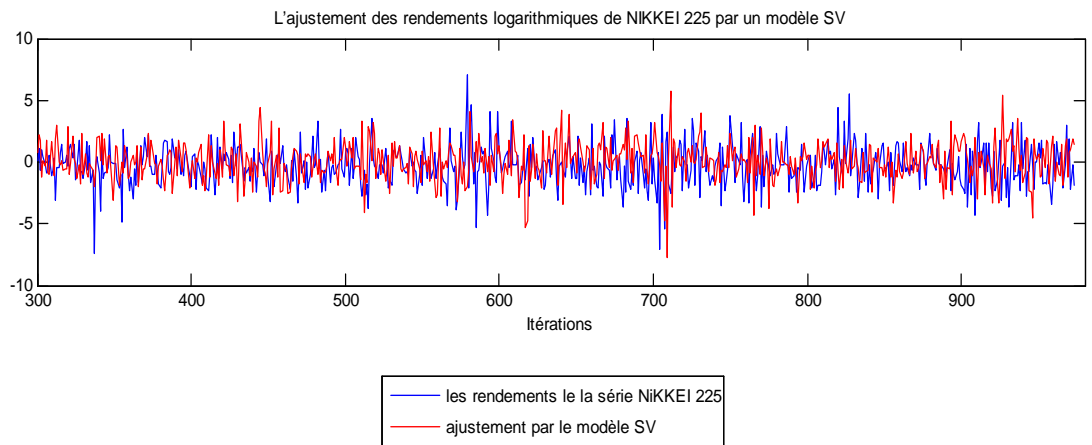
$$h_{t+1} = 0.7027 + 0.9463 (h_t - 0.7027) + 0.1706 \eta_t$$

$$h_1 \rightarrow N \left(0.7027, \frac{0.1706^2}{1 - 0.9463^2} \right)$$

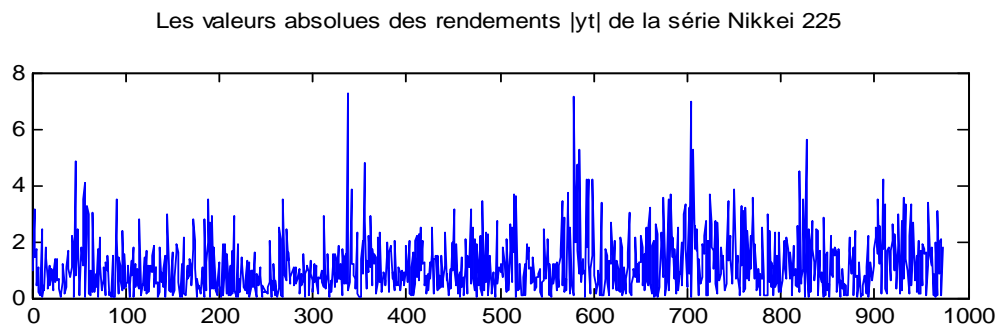
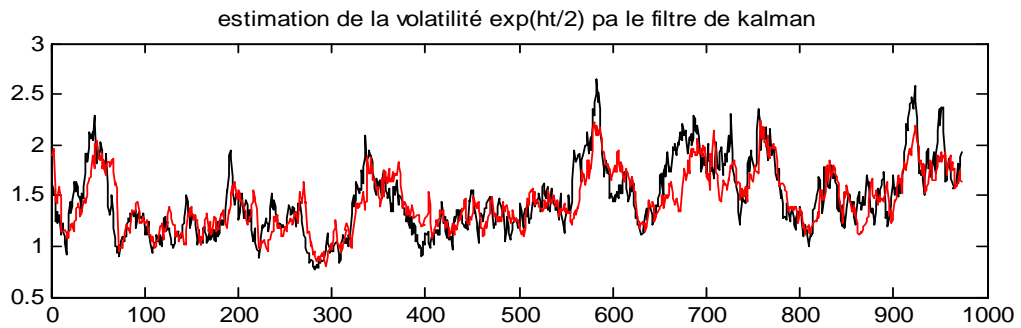
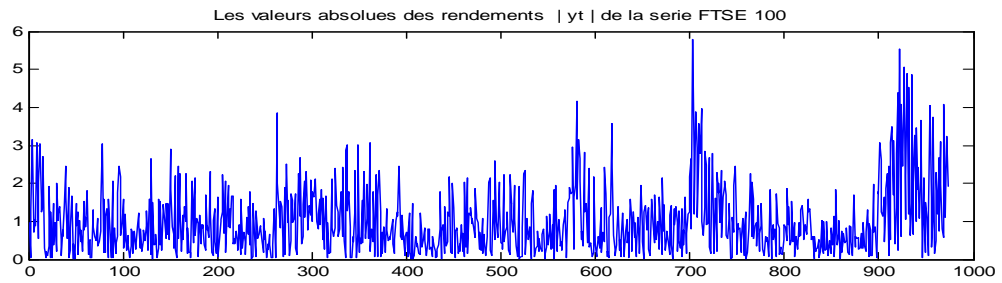
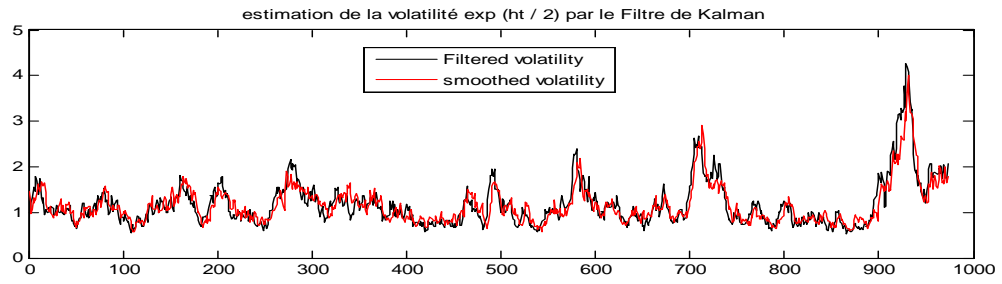
$$\varepsilon_t \rightarrow N(0, 1)$$

$$\eta_t \rightarrow N(0.1)$$

Les figures sont une représentation conjointe de la série des rendements logarithmiques et de la série ajustée et les résidus correspondant.



Les résultats de l'estimations de volatilité filtrées et lissées sont donnés à la figure suivante, avec un graphique des valeurs absolues des rendements. Le graphique suivant montre la fonctionnalité attendue de la volatilité filtrée et la volatilité lissée. Tout au long de l'échantillon, la volatilité filtré est légèrement plus élevé que les valeurs lissées en raison de la baisse progressive de la volatilité observée pour les séries FTSE 100 et Nikkie 225.



Chapitre 5

Modèle à volatilité stochastique périodique

5.1 Introduction

Au cours des dernières décennies, de nombreuses études menées sur des séries financières ont focalisé sur "l'effet saisonnalité". Une liste non exhaustive comprend les travaux d'*Abraham et Ikenberry (1994)*, *Bessenmbinder et Hertznel (1993)*, *French (1980)*... , Ces études se sont souvent intéressées à " l'effet week end", qui pour un grand nombre de place bourcières se traduit par par des rendements significativement plus faible, voir même négatifs les luindi et des rendements souvent fort élevés le vendredi avant la clôture des marchés.

Nonobstant à la volatilité des séries, *Abraham et Ikenberry (1994)* et *Bessenmbinder et Hertznel (1993)* utilisèrent des Autorégressions périodiques pour l'analyse de la structure d'autocorrélation des rendements aux alentours des jours de fermeture. Pourtant, si on considère, maintenant, la variabilité instantanée des séries, des observations similaires peuvent être faites, puisque la volatilité du marché semble en général être plus importante le lundi que le reste de la semaine. Ainsi, la structure dynamique de la volatilité parraît elle aussi témoigner d'assez évidents "effet saisonniers" (*cf. Ederington et Lee (1993)*)

Dans ce chapitre, on s'attellera à présenter la modélisation des processus périodiquement corrélés. Tout d'abord, on exposera les définitions des processus périodiquement corrélés, puis on donnera un aperçu succinct des modèle à volatilité stochastique à coefficients périodiques

5.2 Processus périodiquement corrélé

5.2.1 Historique :

La moitié de siècle, a vu l'émergence de diverse méthodes décrivant les caractéristiques stochastique de processus de séries chronologiques trouvant leurs applications dans plusieurs domaines tels que l'économie, l'environnement, et autres.

En effet depuis les travaux de *Wold (1938)*, de nombreux chercheurs se sont attelés sur cette voie et l'an 1970, fût, à travers le fameux ouvrage de *Box et Jenkins (1976)*, la consécration de tant d'années de labeur et de recherche sur des méthodes d'analyse de séries chronologiques.

Connue dans la littérature sous le nom de *Méthodologie de Box et Jenkins*, ce recueil est devenu, un livre référence sur le traitement, selon une approche standard, des modèle autorégressif-moyenne mobile univariés à coefficients constants, d'où l'appellation anglo-saxonne de «*standard text book*» selon *Ghysels et al (1996)*. Cependant, de nombreux séries rencontrées dans la pratique, font apparaître une tendance saisonnière, pour laquelle des transformations adéquates (différence ordinaire, différence saisonnière, différence mixte, ...) sont usuellement appliquées afin d'assurer la stationnarité au seconde ordre. Toutefois, dans l'analyse des séries exhibant une saisonnalité forte ou une tendance cyclique, supposé une homogénéité au seconde ordre et quelle fois, clairement, non appropriée.

A travers l'exemple de *McCollister et Wilson (1775)*, concernant la mesure heure par heure du taux d'ozone ambiant du bassin de Los Angeles (USA), durant plusieurs mois, *Tiao et Grupe (1980)*, ont mis en évidence les conséquence d'une analyse standard.. En effet une telle approche, a conduit à une mauvaise spécification du modèle homogène car elle n'as pas pris en considération la nature périodique des données. De même *Osborn (1992)* a montré qu'une telle modélisation, erronée, conduit à une surévaluation des ordres, la non-

convergence des estimateurs des paramètres et l'inéfficacité de la prévision. dès lors, cela devenait plus approprié et plus judicieux de traiter les processus périodiquement corrélés, non pas par la classe des modèles saisonniers classiques, mais plutôt par celle des modèles linéaires périodiques autoregressifs moyenne mobile $PARMA_S(p_t, q_t)$ dont les paramètres $\phi_{t,i}$ $i = 1 \dots p_t$ et $\theta_{t,i}$ $i = 1 \dots q_t$, les ordres, ainsi que la variance du processus bruit blanc, sont des fonctions périodiques du temps, de période S , $S \geq 2$

Dans le célèbre théorème de décomposition linéaire de *Cramér (1961)*, connu sous le nom *Décomposition de Wold-Cramér* qui n'est qu'une généralisation de théorème de *Wold (1983)*, les modèles linéaires périodiques, une classe particulière des modèles évolutifs dans le temps, s'avèrent inadéquats pour la représentation des processus périodiquement corrélés. originellement présentés par *Gladyshev (1961)* afin de représenter les processus, du second ordre, dont les fonctions d'autocovariance sont des fonctions périodiques du temps, de période S , les processus périodiquement corrélés, connurent un essor fulgurant, et cela grâce au lien établi avec les processus multivariés stationnaires, qui leur sont associés [*Pagano, M. (1978)*, *Newton, H. J. (1982)*].

En effet *Gladyshev (1961)* donna une définition formelle de la stationnarité périodique des processus périodiques, et montra qu'une condition nécessaire et suffisante, pour une telle stationnarité, est la stationnarité du vecteur contenant toutes les variables périodiques.

Une approche connue sous "period span lumping" ? a fait l'objet de nombreux travaux : *Ula et Smadi (1997)*, *Franses (1996)*, *Bentarzi (1995)*, *Bentarzi et Hallin (1994)*, *Tiao et Grupe (1980)*, *Cleveland et Tiao (1979)*, ces deux dernières publications furent les pionnières, puisqu'elles donnèrent les formules relatives aux paramètres, et à l'ordre du modèle $PARMA_S$ à ceux des processus $ARMA$ S -varié correspondant.

Certains aspects de l'étude des modèles de séries chronologiques périodiques restent difficiles à analyser par le biais de cette approche, on cite à titre d'exemple, le problème de la factorisation spectrale d'un processus périodiquement corrélé q dépendant univarié ou multivarié et les tests (*Bentarzi (1995) et (1998)*, *Bentarzi et Hallin (1994)*, *(1996) et (1998)*).

5.2.2 Processus périodiquement corrélé

Soit y_t un processus du second ordre, de moyenne :

$$\mu_t = E(y_t), \quad t \in \mathbb{Z} \quad (5.1)$$

et de fonction d'autocovariance :

$$\gamma(t, s) = Cov(y_t, y_s), \quad t, s \in \mathbb{Z} \quad (5.2)$$

Sans perte de généralité on suppose que $\mu_t = 0 \forall t \in \mathbb{Z}$, i.e on assume que y_t est un processus centré.

Un cas particulier de cette classe de modèle, est le processus périodiquement corrélés, introduit initialement par *Gladyshev (1961)*.

Définition 5.1

Un processus du second ordre $\{y_t, t \in \mathbb{Z}\}$ est dit Périodiquement Corrélé ou faiblement Périodique s'il existe un entier positif S tel que $\forall t, s$ et $\tau \in \mathbb{Z}$, on a :

$$\mu_{t+S\tau} = \mu_t, \quad (5.3.a)$$

Et

$$\gamma(t + S\tau, s + S\tau) = \gamma(t, s), \quad (5.3.b)$$

L'entier positif S est dans ce cas est appelé la Période du processus périodiquement corrélé.

On remarque qu'un processus faiblement stationnaire de second ordre est un cas particulier des processus périodiquement corrélés. ($S = 1$)

5.2.3 Fonction d'autocovariance d'un processus périodiquement corrélé

On considère un processus y_t centré, périodiquement corrélé, de période S .

Soient, pour tout $t \in \mathbb{Z}$ l'entier positif i , $i = 1, \dots, S$ et l'entier $\tau \in \mathbb{Z}$, tels que :

$$t = i + S\tau \quad (5.4)$$

Dés lors la fonction d'autocovariance, d'ordre h relative à la $i^{\text{ème}}$ période du processus S périodique y_t noté $\gamma_h^{(i)}$ est définie comme suit : (voir *Bentarzi (1995)* et *Franses(1996)*)

$$\gamma_h^{(i)} = \gamma(t, t - h) = \gamma(i, i - h), \quad i = 1, \dots, S \text{ et } t \in \mathbb{Z} \quad (5.5)$$

Il est facile de vérifier que :

$$\gamma_h^{(i+S\tau)} = \gamma_h^{(i)} \quad \text{et} \quad \gamma_{-h}^{(i)} = \gamma_h^{(i+h)} \quad (5.6)$$

5.2.4 Processus périodiquement corrélés et processus multivariés stationnaires

Processus périodiquement corrélés et processus multivariés stationnaires de prime abord, il semble qu'il n'y a aucun lien entre eux, ce constat est erroné, en effet dès 1961, Gladyshev établit la relation et a montré que l'on peut représenter un processus périodique, de période S ($S \geq 2$), par un processus S -varié stationnaire. Le modèle ainsi obtenu, est exploité, afin de déduire les propriétés du modèle périodique, i.e. que l'étude des modèles périodiques se ramène à celle du modèles multivariés stationnaire associés.

sois le processus S -varié, $X(\tau)$, dont la $j^{\text{ème}}$ composante, $X_j(\tau)$, est définie par :

$$X_j(\tau) = y_{j+S\tau}, \quad j = 1, 2, \dots, S \text{ et } \tau \in \mathbb{Z} \quad (5.7)$$

Ce processus est alors, donné par le vecteur :

$$\begin{aligned} X(\tau) &= (X_1(\tau), X_2(\tau), \dots, X_S(\tau))' \\ &= (y_{1+S\tau}, y_{2+S\tau}, \dots, y_{S+S\tau})' \end{aligned} \quad (5.8)$$

Théorème de Gladyshev (1961)

Un processus y_t est Périodiquement Corrélé de Période S si, et seulement si, le processus $X(\tau)$ S -varié qui lui associé, est faiblement stationnaire.

5.2.5 Relation entre les autocovariances d'un processus périodiquement corrélé et celles processus multivariés stationnaires inhérent

soit $\Gamma(\tau_1, \tau_2)$ la matrice de variance covariances de $X(\tau_1)$ et $X(\tau_2)$, dont les éléments sont donnés par :

$$\Gamma_{ij}(\tau_1, \tau_2) = (X_i(\tau_1), X_j(\tau_2)) \quad i, j = 1, 2, \dots, S \text{ et } \tau_1, \tau_2 \in \mathbb{Z} \quad (5.9)$$

On peut vérifier que le terme (1.9) s'exprime en fonction de la fonction d'autocovariance $\gamma_h^{(i)}$ du processus périodique y_t , comme suit :

$$\Gamma_{ij}(\tau_1, \tau_2) = \gamma_{i-j+S(\tau_1-\tau_2)}^{(i)} \quad i, j = 1, 2, \dots, S \text{ et } \tau_1, \tau_2 \in \mathbb{Z} \quad (5.10)$$

Le processus S-varié $X(\tau)$, étant stationnaire, il en dérive que sa matrice d'autocovariance à l'horizon h est de la forme :

$$\Gamma(h) = \Gamma(\tau, \tau - h) = (\Gamma_{ij}(\tau, \tau - h))_{i,j=1,\dots,S} \quad (5.11)$$

Où $\Gamma_{ij}(\tau, \tau - h) = \gamma_{i-j+Sh}^{(i)} \quad \tau \in \mathbb{Z}$

5.3 Présentation du modèle

Définition :

Un processus aléatoire y_t est dit satisfait un modèle à volatilité stochastique à coefficients périodiques, en t de période S , noté PSV, si il est solution d'une équation aux différences stochastiques de la forme suivante :

$$y_t = \beta_t e^{h_t/2} \varepsilon_t, \quad (5.12.a)$$

$$\begin{aligned} h_{t+1} &= \mu_t + \phi_t (h_t - \mu_t) + \sigma_{\eta,t} \eta_t \\ h_s &\rightarrow N \left(\mu_s, \frac{\sigma_{\eta,s}^2}{1 - \prod_{i=1}^S \phi_i^2} \right) \end{aligned} \quad (5.12.b)$$

Où η_t et ε_t sont des bruits blancs Gaussiens qui sont mutuellement indépendants centrés et de variance égale à 1.

on suppose y_t la serie temporelle, soit observable. Cette variable represente le taux de rendement d'un titre financier, elle depend de la variable h_t qui est pas observable. cette deriere represente le log-volatilité à l'instant t

Les coefficients du système sont des fonctions périodiques du temps, de période S c'est à dire :

$$\begin{aligned}\beta_t &= \beta_{s+S\tau} = \beta_s \\ \phi_t &= \phi_{s+S\tau} = \phi_s \\ \mu_t &= \mu_{s+S\tau} = \mu_s \\ \sigma_{\eta,t}^2 &= \sigma_{\eta,s+S\tau}^2 = \sigma_{\eta,s}^2\end{aligned}$$

Tels que : $\tau, t \in Z$ et $s \in \{1, 2, \dots, S\}$

Simulation

Le processus générateur de donnée et le trajectoire sont données comme suit :

$$\begin{aligned}y_{s+S\tau} &= \beta_{s+S\tau} e^{h_{s+S\tau}/2} \varepsilon_{s+S\tau} , \\ h_{s+S\tau} &= \mu_s + \phi_s (h_{s+S\tau-1} - \mu_s) + \sigma_{\eta,s} \eta_{s+S\tau-1}\end{aligned}$$

Le processus est simulé d'un modèle à volatilité stochastique à coefficients périodiques, PSV de période $S = 2$

Les vecteurs du paramètres périodiques sont : $\underline{\mu} = (0, -0.8)'$, $\underline{\phi} = (0.95, 0.97)'$ et $\underline{\sigma}_{\eta_i} = (0.1, 0.2)'$.

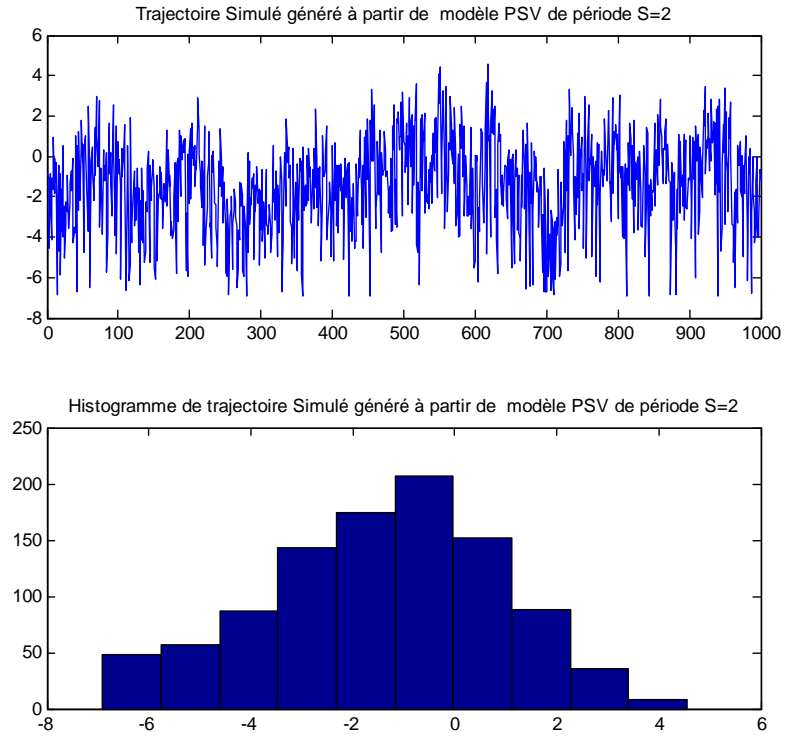
On prend

$$\underline{\beta} = \exp(\underline{\mu}/2) \text{ dans ce cas on trouve } \underline{\beta} = (1, 0.6703)'$$

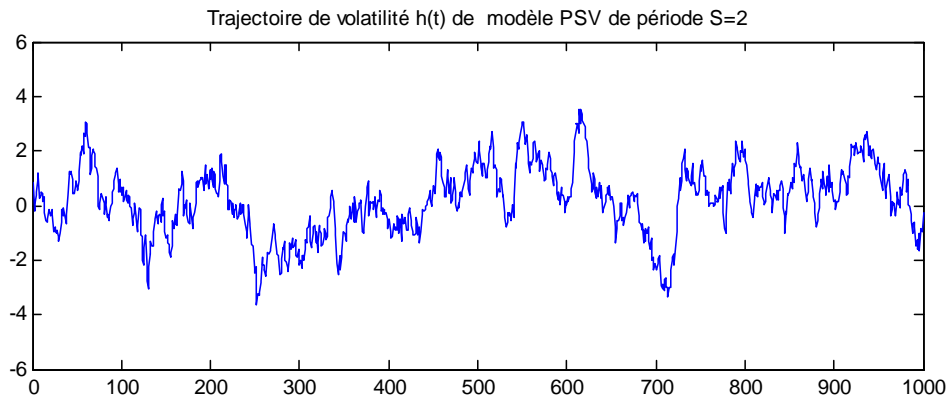
D'où

$$\begin{aligned}y_{1+2\tau} &= \beta_{1+2\tau} e^{h_{1+2\tau}/2} \varepsilon_{1+2\tau} \\ h_{1+2\tau} &= \mu_1 + \phi_1 (h_{1+2\tau-1} - \mu_1) + \sigma_{\eta,1} \eta_{1+2\tau-1}\end{aligned}$$

$$\begin{aligned}y_{2+2\tau} &= \beta_{2+2\tau} e^{h_{2+2\tau}/2} \varepsilon_{2+2\tau} \\ h_{2+2\tau} &= \mu_2 + \phi_2 (h_{2+2\tau-1} - \mu_2) + \sigma_{\eta,2} \eta_{2+2\tau-1}\end{aligned}$$



La figure suivante montre le trajectoire du *volatilité* (h_t)



Chapitre 6

Conclusion

Dans ce travail de mémoire, nous avons considéré le problème de la modélisation de la volatilité des séries financières, en particulier les modèles de séries chronologique à volatilité stochastique.

Nous avons utilisé l'algorithme du Filtre de Kalman pour intégrer l'état non observable. La performance de cette méthode a été montrée par des simulations intensives. D'autre part, nous avons choisi l'approche Bayésienne pour estimer les paramètres du modèle étudié où chaque paramètre est considéré comme étant une variable aléatoire ; pour cela, nous avons appliqué la méthode MCMC. Par la suite, nous avons effectué des applications sur des séries réelles

A travers une série de simulations, nous avons constaté que les estimations obtenues sont efficaces. Nous avons relevé que la méthode MCMC est nettement meilleure, mais son implémentation est lourde ; et par comparaison avec des résultats obtenus par d'autres chercheurs sur des données réelles les nôtres sont considérés bons.

Nous envisageons, suite à ces concluants résultats et en égard à leur importance dans le monde de la finance, d'étendre notre horizon vers l'étude des propriétés probabilistes et statistiques que nous avons commencé dans le chapitre " *Modèle à volatilité stochastique périodique*", tout en oeuvrant à approfondir et enrichir les notions acquises.

Bibliographie

- [1] . Abraham, A. and Ikenberry, D. L. (1994). The individual investor and the weekend effect. *Journal of Financial and Quantitative Analysis*, **29**, 263-78.
- [2] . Andersen, T. G., Bollerslev, T., Diebold, F.X., et Labys, P. (2003). Modeling and Forecasting Realized Volatility, *Econometrica* 71 : 579–625.
- [3] . Baillie, R.T. and R.P. DeGennaro (1990), “Stock Returns and Volatility”, *Journal of Financial and Quantitative Analysis*, Vol. 25, pp. 203-214.
- [4] . Bentarzi, M. and Hallin, M. (1994). On the Invertibility of Periodic Moving Average Models. *Journal of Time Series Analysis*, **15**, 263-268.
- [5] . Bentarzi, M. and Hallin, M. (1996). Locally optimal tests against periodic autoregression. *Econometric Theory*, **12**, 88-112.
- [6] . Bentarzi, M. and Hallin, M. (1998). Spectral factorization of periodically correlated *MA*(1) processes. *Journal of Applied Probability*, **35**, 46-54.
- [7] . Bentarzi, M. (1995), Modèles de séries chronologique a coefficient périodiques, Thèse de doctorat Es-sciences, Institut de Mathématique, U.S.T.H.B., Alger, Algérie.
- [8] . Black, F. (1976). Studies of stock price volatility changes, Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section 177–181.
- [9] . Bollerslev, T.P. (1986) : "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 309–328.
- [10] . Bollerslev, Tim (1987). "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return". *The Review of Economics and Statistics* 69 (3) : 542–547.

- [11] . Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control. San Francisco : Holden-Day.
- [12] . Bresson G, Pirotte A. Économétrie des séries temporelles. Théorie et applications 1st ed. Paris : Presses universitaires de France, 1995.
- [13] . Carnero, A., Pena, D. and Ruiz, E. (2003). Persistence and curtosis in autoregressive stochastic volatility models (manuscript, Universidad Carlos III de Madrid).
- [14] . Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541–53.
- [15] . Chen M.-H., Shao Q.-M. et Ibrahim J. G. (2000). Monte Carlo methods in Bayesian computation. Springer Series in Statistics. Springer-Verlag.
- [16] . Chib, S. (1995). Marginal likelihood from Gibbs output. *J. Am. Statist. Assoc.* 90, 1313–21.
- [17] . Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors. *J. Econometrics* 64, 183–206.
- [18] . Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–35.
- [19] . Chib, S. and E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409–31.
- [20] . Chib, S., F. Nardari, and N. Shephard, (2002). “Markov Chain Monte Carlo Methods for Stochastic Volatility Models,” *Journal of Econometrics*,108, 281-316.
- [21] . Chou, R. Y. (1988), “Volatility Persistence and Stock Valuations : Some Empirical Evidence Using GARCH”, *Journal of Applied Econometrics*, 3, 279-294.
- [22] . Cleveland, W. P and Tiao, G. C. (1979). Modeling seasonal time series. *Revue Economique Appliquée*, **32**, 107-129.
- [23] . Celeux G., Diebolt J. (1985), The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Computational Statistics Quaterly*, 2, 73-82.
- [24] . Cowles M. K. et Carlin B. P. (1996). Markov chain Monte Carlo convergence diagnostics : a comparative review. *J. Amer. Statist. Assoc.*, vol. 91, n434. pages 883–904.

- [25] . Cramér, H. (1961). On some classes of nonstationary stochastic processes. Proceeding of the 4th Berkley Symposium on Mathematical Statistics and Probability, 2, 57-78.
- [26] . Christie, A.A. (1982) : "The Stochastic Behavior of Common Stock Variances : Value, Leverage and Interest Rate Effects" Journal of Financial Economics, 10, 407–432.
- [27] . Danielsson, J. (1994) : "Stochastic Volatility in Asset Prices, Estimation with Simulated Maximum Likelihood," Journal of Econometrics, 64, 375
- [28] . Dempster .A , Laird. N, et Rubin .D, (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B, 39 :1–38.
- [29] . De Jong, P. (1991). The diffuse Kalman filter. Ann. Statist. 19, 1073–1083.
- [30] . De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. Biometrika 82, 339–50.
- [31] . Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.
- [32] . Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). Journal of the Royal Statistical Society B 62 : 3–56.
- [33] . Durham G. B., (2006) ,Monte Carlo methods forestimating, smoting, and filtering one- and two-factor stochastic viltility models, Journal of Econometrics.
- [34] . d'Estampes L., Garel B. et Saint Pierre G. (2003). Test séquentiel : Niveau de confiance après acceptation. Revue de Statistiques Appliquées, A paraître.
- [35] . Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica 50, 987–1007.
- [36] . Engle, R., V. Ng. 1993. Measuring and testing the impact of news on volatility. J. Finance 48 1749–1778.
- [37] . Engle, R.,, 1983, Estimates of the variance of U.S. inflation based on the ARCH model, Journalof Money Credit and Banking 15, 286-301.
- [38] . Engle, R. and D. Kraft, 1983, Multiperiod forecast error variances of inflation estimated fromARCH models, in : A. ZeUner, ed., Applied time series analysis of economic data (Bureau of the Census, Washington, DC) 293-302.

- [39] . Fama, E., (1965), "The Behaviour of Stock Market Prices", *Journal of Business* 38, pp. 34-105.
- [40] . Francq C. et Zakoian. J.M (2009) Modèles GARCH et à volatilité stochastique. Support de cours, [http ://perso.univ-lille3.fr/~cfrancq](http://perso.univ-lille3.fr/~cfrancq)
- [41] . Franses, P. H. (1996). *Periodicity and Stochastic Trends in Economic Times Series*. Oxford University Press.
- [42] . French, K. R., G. W. Schwert and R. E. Stambaugh (1987), "Expected Stock Returns and Volatility", *Journal of Financial Economics*, 19, 3-29.
- [43] . Gallant, Ronald & Tauchen, George, 1989. "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes : Asset Pricing Applications," *Econometrica*, Econometric Society, vol. 57(5), pages 1091-1120, September
- [44] . Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions*, PAMI 6, 721–41.
- [45] . Gentle J.E.(2002). *Elements of computational statistics*. Statistics and Computing. SpringerVerlag.
- [46] . Geweke, J. (1994). Comment on bayesian analysis of stochastic volatility models. *J. Business and Economic Statist.* 12, 397–399.
- [47] . Ghysels, E., A.C.Harvey, and E.Renault (1996). Stochastic volatility. In C.R.Rao and G.S. Maddala (Eds.), *Statistical Methods in Finance*, pp.119–191. Amsterdam : North-Holland.
- [48] . Ghysels,E. et A Hull et H. S. Lee (1996).On periodic structures and testing for seasonal unit roots, *.Journal of the American Statistical Association*, 91, 436 : 1551-59
- [49] . Gladyshev.(1961) Periodically random sequences.*Soviet Mathematics*,2 :385-388,1961.
- [50] . Glosten, Lawrence R & Jagannathan, Ravi & Runkle, David E, 1993. " On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, American Finance Association, vol. 48(5), pages 1779-1801, December
- [51] . Green P. (1995). Reversible jump mcmc computation and bayesian model determination. *Biometrika*, vol. 82. pages 711–732.

- [52] . Gilks, W. K., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London : Chapman & Hall.
- [53] . Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge : Cambridge University Press.
- [54] . Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Rev. Economic Studies* 61, 247–64.
- [55] . Hautsch N et Ou Y.(2008), *Discrete-Time Stochastic Volatility Models and MCMC-Based Statistical Inference*, Humboldt University, Collaborative Research Center 649
- [56] . Hentschel, L. 1995. All in the family : Nesting symmetric and asymmetric GARCH models. *J. Financial Econom.* 39 71–104.
- [57] . Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* 42, 281–300.
- [58] . Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Business and Economic Statist.* 12, 371–417.
- [59] . Jacquier, E., N. G. Polson, and P. E. Rossi (1995). Models and prior distributions for multivariate stochastic volatility. Unpublished paper : Graduate School of Business, University
- [60] . J.McCausland, *Drawing Stochastic Volatility*, Université de Montréal, CIREQ and CIRANO, Mars 2008
- [61] . Kim, S, Shephard, N. and Chib, S. (1998), "Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, 45, 361-393.
- [62] . Krichene, N. 2003, *Modeling stochastic volatility with Application to stock return*. IMF International Monetary Fund, working paper, WP 03/125.
- [63] . Mandelbrot, B. (1963) : "The Variations of Certain Speculative Prices," *Journal of Business*, 36, 394–419.
- [64] . Manabu Asai, (2009) Bayesian analysis of stochastic volatility models with mixture-of-normal distributions, *Math. Comput. Simul.* doi :10.1016/j.matcom.2008.12.013.
- [65] . Markowitz, H.M. (March 1952). "Portfolio Selection". *The Journal of Finance* 7 (1) : 77–91. doi :10.2307/2975974. JSTOR 2975974.

- [66] . Melino, A. & Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* 45 : 239–265.
- [67] . McCollister et Wilson (1975), Linear stochastic models for forecasting daily maxima and hourly concentration of air pollutants, *Atmos. Environ* , 9 : 417-23.
- [68] . McLachlan G.J.et Bashford K.E. (1988) Mixture models : inferences and applications to clustering, Marcel Dekker, New York.
- [69] . McLachlan G.J.et Krishnan T. (1997), The EM algorithm and extensions, John Wiley & Sons, New York.
- [70] . McLachlan G.J.et Peel D. (2000), Finite mixture models, John Wiley & Sons, New York.
- [71] . Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45 : 239–265.
- [72] . Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *J Chemical Physics* 21, 1087–92.
- [73] . Nelson D.B. (1991) : "Conditional Heteroskedasticity in Asset Returns : a New Approach," *Econometrica*, 59, 347–370.
- [74] . Newton, H. J.(1982). Using periodic autoregressions for multiple spectral estimation. *Technometrics*, 6, 1310-1317
- [75] . Osborn, D. R. (1992).The implication of periodically varying coefficients for seasonal time-series processes. *Journal of Econometrics*, 48, 373-384.
- [76] . Pitt, M. K. and N. Shephard (1997). Filtering via simulation : an auxiliary variable approach. working paper, Nuffield College, Oxford.
- [77] . Pitt, M. and N. Shephard (1999). Filtering via simulation : auxiliary particle filter. *Journal of the American Statistical Association* 94, 590–599.
- [78] . Pagano, M. (1978). On periodic and multiple autoregression. *The Annals of Statistics*, 6, 1310-1317.
- [79] . Poon, S., and S. J. Taylor (1991), “Macroeconomic factors and the UK stock market”, *Journal of Business Finance and Accounting* 18, pp. 619-636.

- [80] . Robert C. P. (1996a). Méthodes de Monte Carlo par chaînes de Markov. Éditions Économica.
- [81] . Robert C. P. et Casella G. (1999). Monte Carlo statistical methods. Springer-Verlag.
- [82] . Ruiz, E. (1994) : Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Econometrics* 63, 289–306.
- [83] . Sandmann, G. & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* 87 : 271-301.
- [84] . Shephard, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *J. Appl. Econometrics* 8, S135–52.
- [85] . Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* 81, 115–31.
- [86] . Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, O. E. Barndorff-Nielsen, and D. V. Hinkley (Eds.), *Time Series Models in Econometrics Finance and Other Fields*, pp. 1–67. London : Chapman & Hall.
- [87] . Shephard, N. and S. Kim (1994). Comment on ‘Bayesian analysis of stochastic volatility models’ by Jacquier, Polson and Rossi. *J. Business and Economic Statist.* 11, 406–410.
- [88] . Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–67.
- [89] . Tauchen, George E & Pitts, Mark, 1983. "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, *Econometric Society*, vol. 51(2), pages 485-505, March.
- [90] . Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79, in O. D. Anderson (ed.), *Time Series Analysis : Theory and Practice* 1, North-Holland, Amsterdam, pp. 203–226.
- [91] . Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester : John Wiley.
- [92] . Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance* 4, 183–204.
- [93] . Tiao, G. C. and Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, **67**, 365-373.

- [94] . Tierney L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.*, vol. 22, n 4. 1701–1762.
- [95] . Tierney L. (1996). Introduction to general state-space Markov chain theory. In *Markov chain Monte Carlo in practice*, pages 59–74. Chapman & Hall.
- [96] . Titterton D.M., Smith A.F.M., Makov U.E. (1985) *Statistical Analysis of Finite Mixture*, John Wiley & Sons, New York.
- [97] . Ula, T. A. and Smadi, A. A. (1997). Periodic stationary conditions for periodic autoregressive moving average processes as eigenvalue problems. *Water Resources Research*, **33**, 1929-1934.
- [98] . Wold, H.(1938). *A study in the stationary time series*. Almqvist and Wiksell, Stockholm (2nd edition : 1954).
- [99] . Wong, C. S. and Li, W. K. (2001). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association* 96, 982 – 95.
- [100] . Zakořian, J.M. (1994) : "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955.
- [101] . Site internet : <http://finance.yahoo.com/q/hp?s=%5EFTSE+Historical+Prices>
<http://fr.finance.yahoo.com/q/hp?s=%5EN225>