

Nabil Saïmi

Estimation de la volatilité et filtrage non linéaire

Mémoire
présenté
à l'Université du Québec à Trois-Rivières
pour l'obtention
du grade de maître ès sciences (M.Sc.)

Département de mathématiques et d'informatique
UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

Table des matières

1	Introduction	4
2	Séries financières et volatilité	8
2.1	Séries financières	8
2.1.1	Cotation quotidienne	8
2.2	Série de rendement	10
2.2.1	Propriétés statistiques des séries de rendement	14
2.2.2	Moments d'échantillonnage	15
2.2.3	Hétéroscédasticité	16
2.2.4	Autocorrélation	18
2.2.5	La volatilité	23
2.2.6	Conclusion du chapitre	24

3 Modèle de Black-Scholes et estimation de la volatilité	25
3.1 Présentation du modèle de Black-Scholes	25
3.2 Importance de la volatilité	28
3.3 La notion d'option	29
3.3.1 Définition	29
3.3.2 Les deux formes d'options	30
3.3.3 Option européenne et option américaine	30
3.3.4 La notion d'arbitrage et la relation de parité call-put .	32
3.4 Estimation de la volatilité à partir de données historiques . . .	33
3.5 La volatilité implicite	38
3.5.1 Définition	38
3.5.2 Modèle d'évaluation	39
3.5.3 Calcul de la volatilité implicite	39
3.5.4 Facteur influençant la volatilité implicite	41
3.5.5 Techniques d'estimation de la volatilité implicite	43
4 Modélisation ARCH	46
4.1 Introduction aux modèles ARCH	46

4.2	Présentation théorique des modèles de types ARCH	47
4.2.1	Processus ARCH	48
4.2.2	Processus GARCH	52
4.3	Extensions du modèles GARCH	54
4.3.1	Effets ARCH et coefficient d'aplatissement	59
4.4	Les processus hétéroscédastiques comme approximations de processus en temps continu	60
4.4.1	Convergence des chaînes de Markov vers des diffusions	60
4.5	Conclusion	62
5	Modèles à Volatilité Stochastique	65
5.1	Introduction	65
5.2	Quelques faits caractéristiques	72
5.2.1	Queues épaisse “(thick tails)”	72
5.2.2	Regroupement de la volatilité “(Volatility clustering)”	73
5.2.3	Effets de puissance “(Leverage effects)”	73
5.2.4	Arrivée d'information	74
5.3	Modèles bivariés à volatilité stochastique	75

6 Méthodes de filtrage non linéaire	77
6.1 Description du problème de filtrage de manière générale	78
6.2 Le modèle	80
6.3 Filtrage discret-discret	81
6.3.1 Filtre de Kalman	81
6.3.2 Méthode du Quasi-Maximum de vraisemblance	84
6.3.3 Méthode de filtrage Non-Linéaire NFML: Non Linear Filtering Maximum Likelihood	87
6.3.4 Comparaison des estimations	91
6.4 Théorie de filtrage continu-discret non linéaire	95
6.5 Estimateur des moments conditionnels	98
6.6 Filtres non linéaires approximatifs	101
6.6.1 Version scalaire du filtre tronqué du second ordre (TSOF)	101
6.6.2 Version scalaire du filtre gaussien du second ordre (GSOF)	107
7 Résultats et conclusions	110
7.1 Discrétisation des équations différentielles stochastiques	110
7.1.1 Méthode d'Euler	111

7.2	Méthode du maximum de vraisemblance	113
7.3	Simulation	115
7.4	Application aux données financières	121
7.5	Conclusion	123
A	Mouvement brownien	131
A.1	Définition du mouvement brownien	131
A.2	Propriétés du mouvement brownien	132
A.3	Intégrale stochastique et calcul d'Itô	133
A.3.1	Définition de l'intégrale stochastique par rapport à une martingale à temps discret	133
A.3.2	Définition de l'intégrale stochastique par rapport au mouvement brownien	134
A.3.3	Calcul d'Itô	136
A.3.4	Équations différentielles stochastiques	137
B	Algorithme de filtrage et d'estimation	139
C	Programmes Matlab	141

Liste des figures

2.1	Cote de fermeture respectifs des titres suivants: Coca-Cola Company, General Motors, Microsoft et Mobil Oil Company entre le 09/09/1991 et le 10/09/2001	9
2.2	Autocorrélations des rendements arithmétiques et géométriques de Coke.	19
2.3	Autocorrélations des rendements arithmétiques et géométriques de GM.	20
2.4	Autocorrélations des rendements arithmétiques et géométriques de Mobil.	21
2.5	Autocorrélations des rendements arithmétiques et géométriques de Microsoft.	22
3.1	Évolution de la valeur d'une option de vente en fonction du prix d'exercice	31

3.2 Évolution de la valeur d'une option d'achat en fonction du prix d'exercice	31
6.1 Solution de l'équation avant de Kolmogorov dans le cas d'un processus de Wiener	98
7.1 Prévisions avec paramètres $\alpha = 0.00198$, $\beta = 0.01$, $\kappa = 0.0071$, $\xi = 0.016$, $\sigma_e = 1.95$, et $\rho = 0.5$	126
7.2 Prévisions du prix de GM et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)	127
7.3 Prévisions du prix de Coca-Cola et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)	128
7.4 Prévisions du prix de Microsoft et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)	129
7.5 Prévisions du prix de Mobil et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)	130

Liste des tableaux

2.1	Moyenne, déviation standard, coefficient d'asymétrie et coefficient d'aplatissement des 4 séries de rendements.	16
2.2	Écarts types $\times 10^{-2}$ des différentes sous-séries de rendement. .	17
3.1	Volatilité historique et erreur standard du titre de Coca-Cola Company en fonction de l'historique (n).	35
3.2	Volatilité historique et erreur standard du titre de General Motors Company en fonction de l'historique (n).	36
3.3	Volatilité historique et erreur standard du titre de Mobil Oil Company en fonction de l'historique (n).	37
3.4	Volatilité historique et erreur standard du titre de Microsoft en fonction de l'historique (n).	38
5.1	Aperçu des différentes modélisation de la volatilité stochastique.	76

6.1 L'estimateur de la moyenne selon Watanabe et le RMSE(Root mean squared error) comparé à d'autres estimateurs [Watanabe, 1999]	95
7.1 Résultats de l'estimation des paramètres du modèle de Black-Scholes-Courtadon de la volatilité stochastique.	119
7.2 Estimation des paramètres des séries des cotes de fermetures des titres de GM, Coca-Cola, Microsoft et Mobil.	123

Dédicace

Je dédie le présent mémoire à mes parents en témoignage de leur soutien et de leurs encouragements tout au long de mes études.

Remerciements

À travers cette page, j'aimerais remercier mon directeur de recherche M. Bruno Remillard pour le support qu'il m'a apporté et la confiance qu'il m'a témoignée au cours des deux dernières années. Il m'a donné le goût de la recherche et il m'a permis par sa rigueur et ses connaissances d'améliorer mes aptitudes dans ce domaine.

Je tiens aussi à remercier le centre d'excellence MITACS (Mathematics of Information Technology and Complex Systems); ainsi que le groupe Lockheed Martin Canada pour leur soutien technique et financier.

Chapitre 1

Introduction

“Suppose we use the standard deviation ...of possible future returns on a stock ...as a measure of its volatility. Is it reasonable to take that volatility as constant over time? I think not. ” Fisher Black.

Le concept de volatilité est probablement un des sujets qui suscitent le plus de recherches dans le domaine de la finance mathématique. Cet intérêt pour la volatilité est motivé par deux raisons importantes: le nombre de plus en plus grand de compagnies utilisant les outils de gestion des risques et le grand nombre de produits dérivés transigés dans les marchés financiers mondiaux. Le prix juste d'une option est déterminé par un certain nombre de facteurs dont la volatilité de l'actif sous-jacent. Tous ces facteurs sont directement observables sur le marché mis à part la volatilité. De même, quand une compagnie veut étudier son degré d'exposition à un risque financier, elle doit être

capable d'évaluer la volatilité de chacun des biens qu'elle possède. L'étude la plus reconnue dans le monde de la finance mathématique est certainement le modèle d'évaluation des options de Black-Scholes (1973), qui a connu un succès immédiat auprès des chercheurs et des professionnels de la finance, tant le besoin était grand d'outil de gestion des risques. Le modèle de Black-Scholes permet d'évaluer le prix d'une option dans le cas de non arbitrage, en se basant sur l'hypothèse que l'évolution d'un titre suit un mouvement brownien géométrique et que la volatilité est constante.

Cependant, l'hypothèse de la constance de la volatilité dans la formule de Black-Scholes était rejetée depuis le début tel que démontré dans les travaux de Black (1976), Schmalensee et Trippi (1978) et ceux de Christie (1982); et surtout après le crash de 1987, qui a mis en évidence l'effet "smile" qui pourtant ne devrait pas apparaître sous les hypothèses de Black-Scholes. Ces constatations nous amènent à considérer la volatilité comme processus de diffusion, ce qui a donné naissance aux modèles à volatilité stochastique. Et comme nous ne disposons pas d'information sur ce processus de diffusion de la volatilité, plusieurs modèles, dérivés du modèle général d'Orstein-Uhlenbeck, ont été proposé dans la littérature tel que nous le présentons au chapitre IV du présent mémoire.

La méthode du maximum de vraisemblance n'est pas facile à implémenter dans le cas d'un modèle à volatilité stochastique, Ghysel (1996) et Shepard (1996). Quelques auteurs ont relié cette méthode à la méthode des moments généralisés (GMM), Melino et Turnbull (1990), Andersen et Sorensen (1996) et à la méthode des moments simulés (SMM), Duffie et Singleton

(1993), Gouriéroux (1993). Une autre approche est celle de la méthode du quasi-maximum de vraisemblance (QML) développé par Nelson (1988), Harvey (1994) et Ruiz (1994). Ils appliquent un filtre de Kalman standard à la représentation linéaire de l'espace d'état du modèle à volatilité stochastique. Cependant, l'évidence Monte-Carlo de Jacquier (1994) démontre que la méthode GMM et QML n'ont pas une grande performance car ils ne dépendent pas de la vraie vraisemblance.

Les méthodes de filtrage non linéaire, Maybeck (1982) et Jazwinski (1970), permettent d'obtenir une estimation de variables aléatoires économiques telle que la volatilité, dans une large classe de modèles en temps continu. L'estimation de la volatilité stochastique et de la variance associée à l'estimée obtenue ont une grande implication sur l'analyse des séries financières, sur la gestion de risque ainsi que sur l'évaluation des produits dérivés. Les méthodes de filtrage non linéaire permettent aussi en appliquant une méthode du maximum de vraisemblance l'estimation des paramètres dans les modèles multivariés à volatilité stochastique en utilisant les observations en temps discret.

Dans le présent mémoire, nous nous intéressons à la modélisation et l'estimation de la volatilité dans le cadre d'une modélisation à volatilité stochastique avec des observations en temps discret. Dans un premier temps, nous présentons au prochain chapitre les caractéristiques des séries financières importantes dans la modélisation de la volatilité, ainsi que l'introduction et la définition de concepts essentiels au sujet du présent mémoire, telles les séries de rendement, l'autocorrélation, le coefficient d'asymétrie et le coefficient d'aplatissement. Au troisième chapitre, nous présentons le modèle de Black-Scholes, la notion

d'option, ainsi que les méthodes traditionnelles d'estimation de la volatilité, à savoir la volatilité historique et implicite. Au chapitre 4, nous présentons les modèles conditionnellement hétéroscédastiques, “*autoregressive conditional heteroscedastic ou ARCH*”, dans le cadre d'une modélisation en temps discret. Au chapitre 5, nous nous intéressons à la modélisation de la volatilité stochastique en se basant sur les faits caractéristiques de la volatilité. Nous passons en revue les différentes modélisations présentes dans la littérature. Au chapitre 6, nous présentons les différentes méthodes d'estimation, ainsi qu'un résumé des méthodes de filtrage utilisées. Le chapitre 7 contient les résultats des simulations du filtre gaussien non linéaire que nous avons développé ainsi que les résultats de l'application aux séries de données financières. Nous conclurons ce dernier chapitre par une discussion des résultats obtenus ainsi qu'une série de recommandations pour les études futures.

Chapitre 2

Séries financières et volatilité

Le but du présent chapitre est la présentation des caractéristiques des séries financières importantes dans la modélisation de la volatilité, ainsi que l'introduction et la définition d'un nombre de concepts essentiels au sujet du présent mémoire.

La section 1.2.1 présente le problème des observations incomplètes. La section 1.2.2 introduit la notion de profit et celle de série de rendement ainsi que quelques unes des caractéristiques communes aux séries de rendement.

La section 1.2.3 définit formellement le concept de volatilité.

2.1 Séries financières

2.1.1 Cotation quotidienne

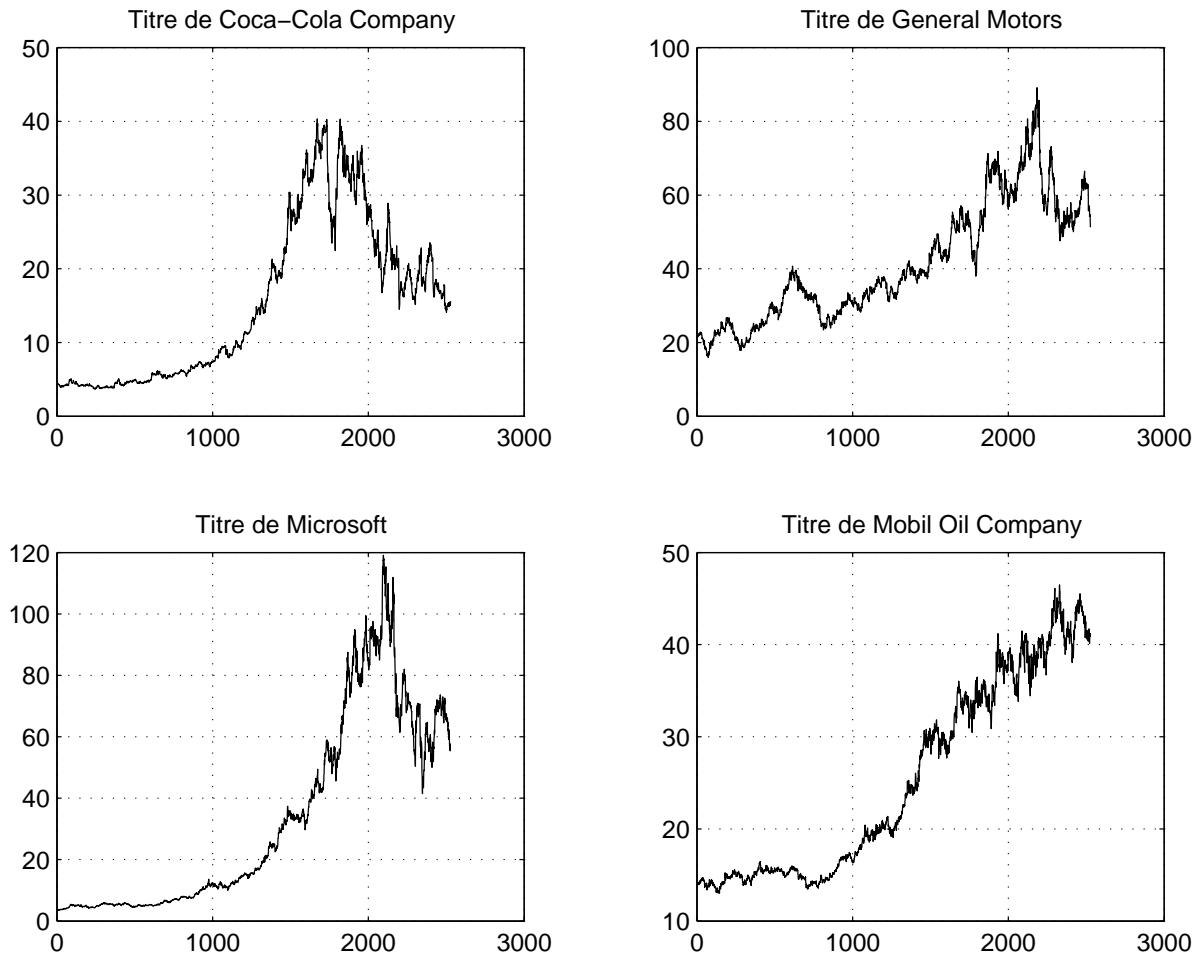


Figure 2.1: Cote de fermeture respectifs des titres suivants: Coca-Cola Company, General Motors, Microsoft et Mobil Oil Company entre le 09/09/1991 et le 10/09/2001

Quatre séries représentant les prix de fermeture quotidiens de quatre titres financiers sont utilisées pour les besoins d'illustration. Il est à noter que ces prix de fermeture “quotidien” ne sont pas vraiment quotidiens dans le sens littéraire du terme. Les marchés boursiers sont fermés pendant les fins de semaine et les jours fériés ce qui implique l'existence de jours non commerciaux, c'est à dire des jours sans cote de fermeture officielle. Dans le présent mémoire, les jours non ouvrables ne seront pas pris en compte dans le sens que chacune des quatre séries sera traitée comme série d'observations équidistantes. Ceci doit être pris en compte lors des résultats statistiques obtenus en utilisant les quatre séries. Les jours de fermeture des marchés boursiers n'ont pas été pris en compte car il n'est pas évident de traiter avec ces observations “manquantes”. Des études de Burghart et Hanweck Jr (1993) indiquent que durant les fins de semaine et les jours fériés, la volatilité est différente de la volatilité des autres jours ouvrables, cependant ils n'ont pas pu quantifier cette volatilité d'une manière qui lui permette d'être incorporée dans les modèles de la volatilité généralement utilisés.

2.2 Série de rendement

En général il n'est pas très intéressant de voir juste le prix d'un investissement, à savoir un titre financier dans notre cas. Du point de vue d'un investisseur, le rendement de l'investissement est beaucoup plus intéressant. Principalement à cause qu'un investisseur insiste plus sur le gain relatif réalisable, plutôt que sur le prix nominal de l'investissement, mais aussi parce que le

rendement comme indice de changement du prix relatif permet des comparaisons entre compagnies, titres boursiers et monnaies. En plus, du fait que les rendements sont généralement stationnaires, une propriété que ne possèdent pas les prix actuels des titres. Dans le monde de la finance, le concept de rendement n'est pas défini de manière claire. Soit S_t le prix d'un titre au temps t . Le rendement à l'instant t peut être défini par :

Définition 2.1 : (Le taux de rendement arithmétique)

$$R_{1,t} = \frac{S_t - S_{t-1}}{S_{t-1}}$$

Une autre définition :

Définition 2.2 : (Le taux de rendement géométrique)

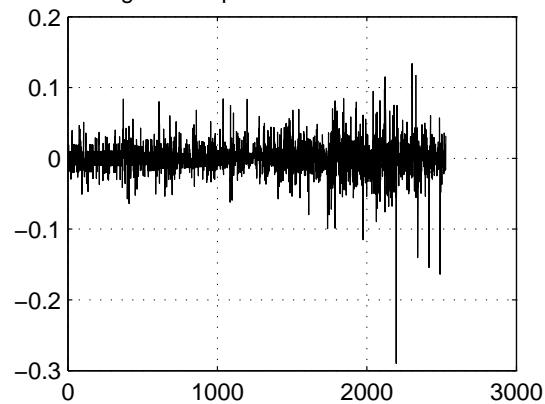
$$R_{2,t} = \log \left(\frac{S_t}{S_{t-1}} \right)$$

Les deux rendements sont liés par la formule suivante, du moins pour les rendements quotidiens:

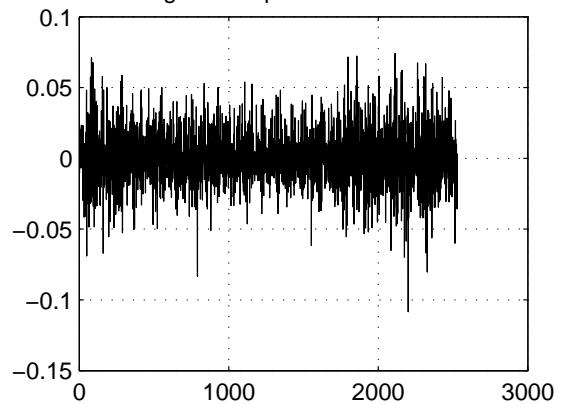
$$R_{1,t} = e^{R_{2,t}} - 1 \approx R_{2,t}$$

Le taux de rendement géométrique est aussi dit rendement composé. Tout au long de ce mémoire, nous utiliserons le taux de rendement géométrique. Il sera noté par R_t et fera référence au rendement. Le taux de rendement géométrique est employé car il est le plus utilisé dans les diverses recherches, ce qui permet de comparer les résultats obtenus, et parce que le taux de rendement géométrique permet de relier les modèles en temps discret et ceux en temps continu.

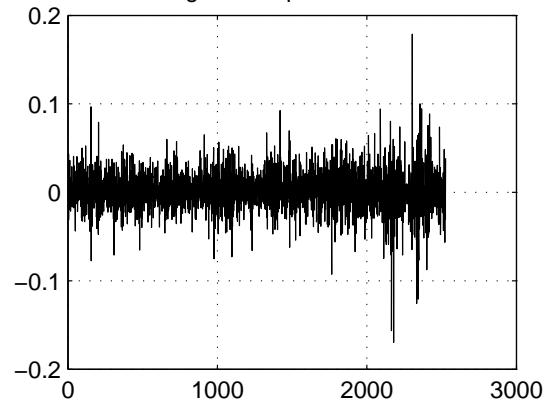
Rendement géométrique du titre de Coca-Cola Company



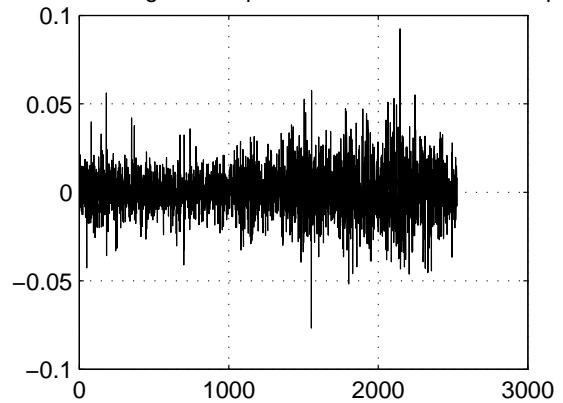
Rendement géométrique du titre de General Motors



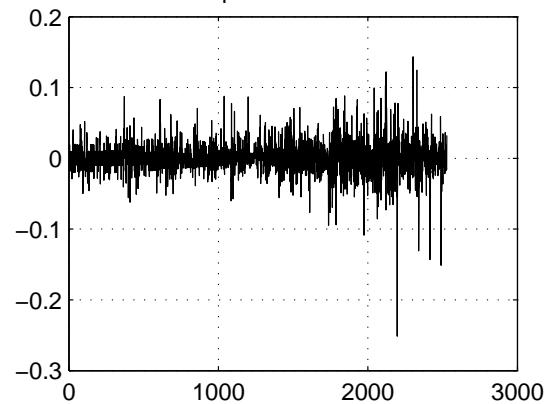
Rendement géométrique du titre de Microsoft



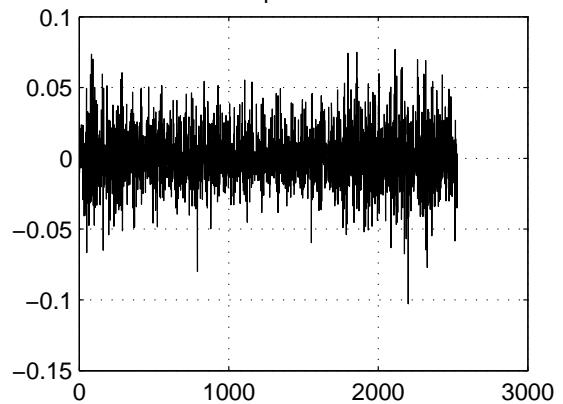
Rendement géométrique du titre de Mobil Oil Company



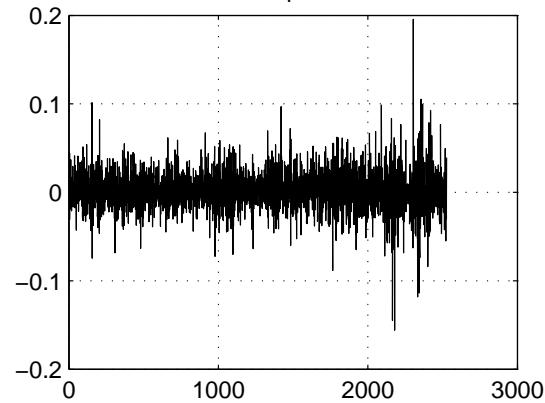
Rendement arithmétique du titre de Coca-Cola Company



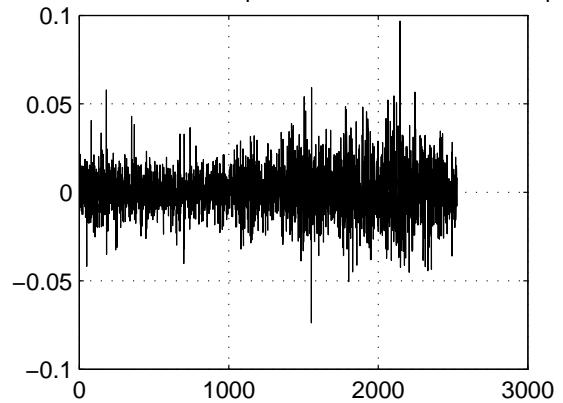
Rendement arithmétique du titre de General Motors



Rendement arithmétique du titre de Microsoft



Rendement arithmétique du titre de Mobil Oil Company



2.2.1 Propriétés statistiques des séries de rendement

En analysant les différentes propriétés statistiques des séries de rendement tels que la moyenne, la déviation standard, l'autocorrélation et les observations extrêmes “points isolés”, il est souvent possible d'avoir une estimation raisonnable du processus générateur de ces séries. Avant, la plupart des modèles représentant les séries de rendement supposaient que les rendements pouvaient être considérés comme étant un bruit blanc gaussien¹. Cependant, nous savons maintenant que cette hypothèse n'est pas correcte. On sait que les séries de rendement possèdent un certain nombre de caractéristiques différentes de celles du bruit blanc gaussien. Les caractéristiques les plus marquantes des séries de rendement sont:

1. **Excès du coefficient d'aplatissement.** Les séries de rendement possèdent un coefficient d'aplatissement plus important que celui du bruit blanc gaussien.
2. **Hetéroscédasticité.** La variance des rendements change dans le temps.
3. **Autocorrélation.** Les rendements ne démontrent qu'une faible autocorrélation, tandis que les carrés des rendements sont autocorrélés à un niveau significatif.

Ces caractéristiques sont généralement connus comme des *Faits de style* que nous développerons plus en détail dans le quatrième chapitre. Les deux

¹C'est à dire une série d'observations indépendantes et identiquement distribuées selon une loi normale.

premiers faits de style ont été décrits dès 1963 par Mandelbrot. Une description plus en détail du troisième fait de style a été faite par Taylor (1994), réclamant que l'autocorrélation des carrés des rendements est positive, statistiquement significative, et généralement plus petite quand le délai entre les observations est plus grand.

2.2.2 Moments d'échantillonnage

Étant données T observations $u_t, t = 1, 2, \dots, T$ d'une certaine variable stochastique, le coefficient d'asymétrie (*skewness*) est calculé par:

$$\frac{E[(U - E(U))^3]}{E[(U - E(U))^3]^{3/2}} = \frac{\frac{1}{T} \sum_{t=1}^T (u_t - \hat{\mu})^3}{(\frac{1}{T} \sum_{t=1}^T (u_t - \hat{\mu})^2)^{3/2}}$$

De la même manière on calcule le coefficient d'aplatissement (*kurtosis*) par:

$$\kappa = \frac{E[(U - E(U))^4]}{E[(U - E(U))^2]^2} \cong \frac{\frac{1}{T} \sum_{t=1}^T (u_t - \hat{\mu})^4}{(\frac{1}{T} \sum_{t=1}^T (u_t - \hat{\mu})^2)^2}$$

où

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T u_t$$

Le coefficient d'asymétrie (*skewness*) est le moment d'ordre 3 divisé par la déviation standard élevée à la puissance 3, tandis que le coefficient d'aplatissement est le moment d'ordre 4 divisé par la déviation standard élevée à la puissance 4. Le tableau suivant résume le calcul de la moyenne, de la déviation standard, du coefficient d'asymétrie et du coefficient d'aplatissement de nos quatre séries de rendements:

Séries	GM	CocaCola	Microsoft	Mobil
Moyenne ($\times 10^{-4}$)	3.482	4.954	-11.242	4.268
Déviation standard	0.0198	0.0227	0.0233	0.0140
Coefficient d'asymétrie	0.0553	-0.9227	0.0917	-0.1521
Coefficient d'aplatissement	1.1270	15.1745	4.448	1.9486

Tableau 2.1: Moyenne, déviation standard, coefficient d'asymétrie et coefficient d'aplatissement des 4 séries de rendements.

2.2.3 Hétéroscédasticité

L'hétéroscédasticité est une des caractéristiques des séries financières, elle représente le caractère non constant de la variance d'une série dans le temps. Quand nous essayons de démontrer l'hétéroscédasticité d'une série, il est plus intéressant de subdiviser notre série de données et ensuite de comparer la déviation standard de chacune des sous-séries obtenues. La table suivante montre la déviation standard pour chacune de nos sous-séries.

En analysant ce tableau, il paraît évident qu'aucune des séries ne possède une variance constante, toutes les séries de rendement démontrent une certaine hétéroscédasticité sur la période d'échantillonnage considérée. La série de rendement sur le titre de Coca-Cola Company est la plus illustratrice. La sous-séries des rendements de l'observation 2100 à 2400 possède une déviation standard qui est le double de la sous-série de rendement allant de l'observation 1 à 300.

Séries	GM	Coke	Microsoft	Mobil
1-300	2.232	1.518	3.207	1.496
301-600	1.780	1.759	3.100	1.958
601-900	1.799	1.715	2.965	1.648
901-1200	1.587	1.790	2.831	1.652
1201-1500	1.503	1.867	2.666	1.241
1501-1800	1.947	2.414	2.565	0.885
1801-2100	2.170	2.809	2.485	0.957
2101-2400	2.646	3.415	2.441	1.132
2401-2527	1.839	2.665	2.436	1.128

Tableau 2.2: Écarts types $\times 10^{-2}$ des différentes sous-séries de rendement.

Il est techniquement difficile de tester l'hétéroscédasticité d'une série quand on ne fait pas de supposition *à priori* sur la distribution de cette série. Si les rendements suivaient une loi normale, on pourrait à ce moment là utiliser le test de Barlett. Mais selon les résultats de la première table, la normalité des rendements est une hypothèse incorrecte. De plus, les résultats d'un tel test seraient difficiles à interpréter dans le cas où l'hypothèse d'homoscédasticité serait acceptée. On ne peut pas savoir si le choix d'un autre pas de division de la série de rendement entraînerait un rejet de l'hypothèse.

2.2.4 Autocorrélation

L'autocorrélation des 4 titres est illustrée par les figures suivantes. Le taux 95% de signification donne $\pm 2/\sqrt{1264} = 0.0563$ comme frontière de rejet.

Il existe des méthodes plus formelles pour exhiber l'existence d'une autocorrélation. Le test de Ljung-Box utilise les valeurs empirique de la fonction d'autocorrélation pour tester l'autocorrélation dans la série de rendement. Si on utilise les rendements au carré, le test est dit test *Portemanteau Q-test*.

Test de Ljung-Box et Portemanteau

Étant donnés des observations $u_t, t = 1, 2, \dots, T$ d'un certain processus stochastique U , le test de Ljung-Box, McLeaod & Li (1983), utilise les estimations empiriques de la fonction d'autocorrélation pour tester l'autocorrélation dans les observations. Si on utilise le carré des observations, on parle à ce moment là du *Portemanteau Q-test* et le test est ensuite utilisé pour tester l'autocorrélation dans les carrés des rendements. Le test suppose que les observations u_t sont indépendantes et identiquement distribuées avec une variance finie.

La statistique de Ljung-Box est définie par:

$$Q(p) = N(N+2) \sum_{k=1}^q \frac{\hat{\rho}^2(k)}{N-k}$$

où l'estimation de la fonction d'autocorrélation est donnée par:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^T (u_t - \hat{\mu})(u_{t-k} - \hat{\mu})}{\sum_{t=1}^T (u_t - \hat{\mu})^2}$$

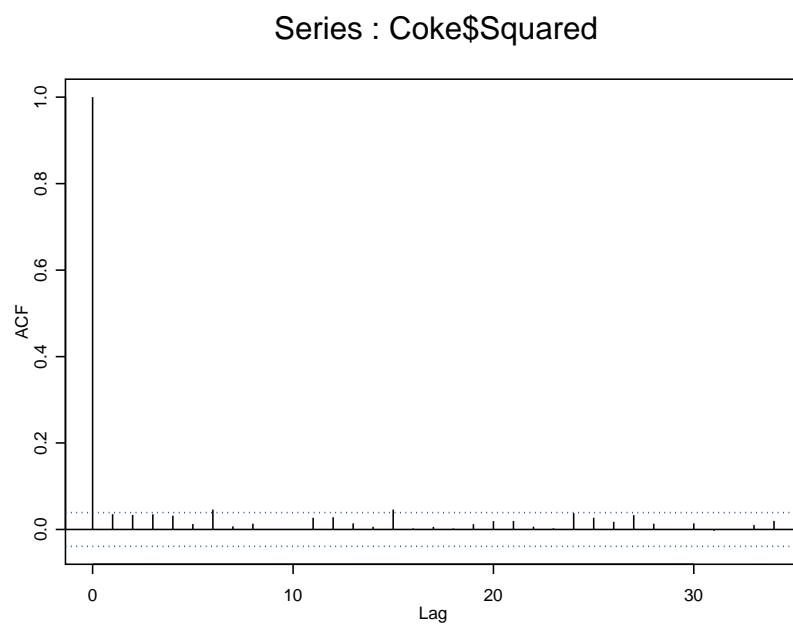
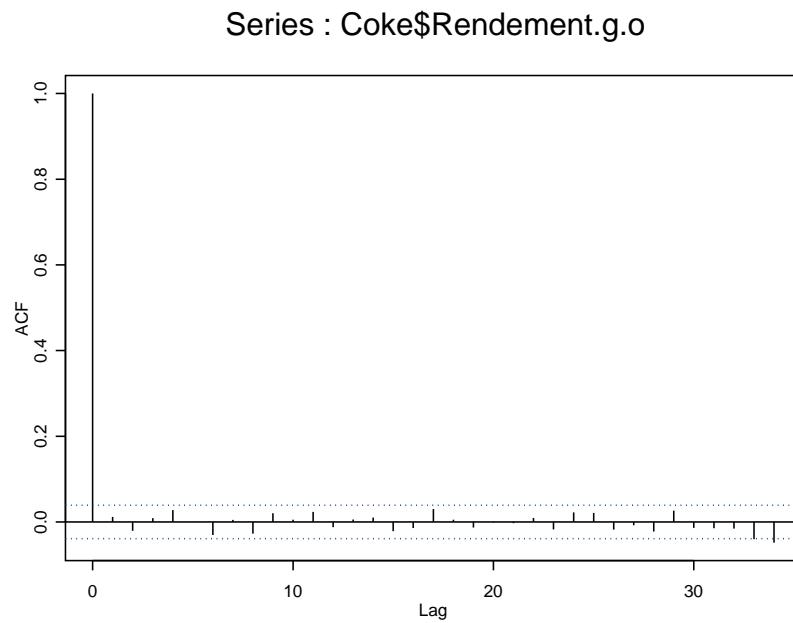
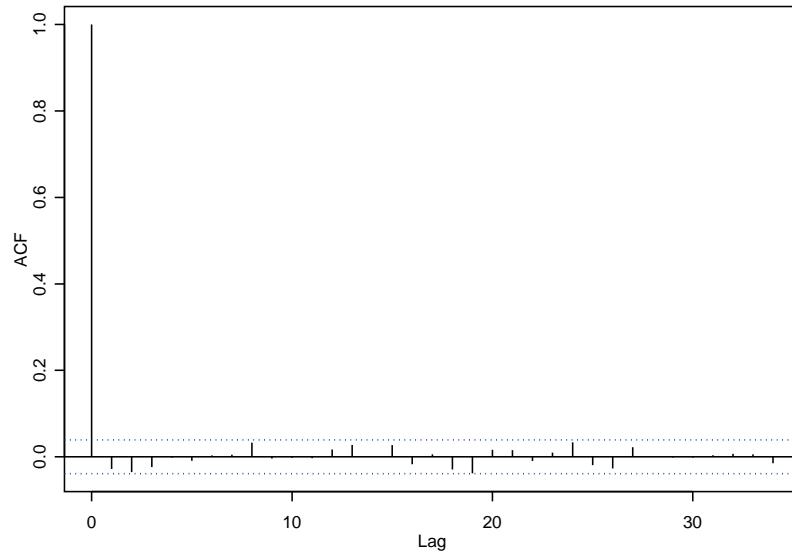


Figure 2.2: Autocorrélations des rendements arithmétiques et géométriques de Coke.

Series : GM1\$Rendement.G.om.trique



Series : GM1\$Squared

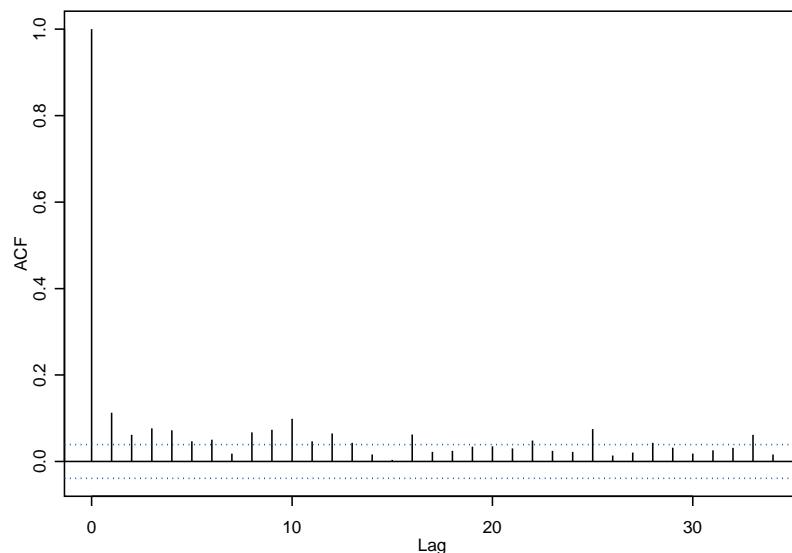
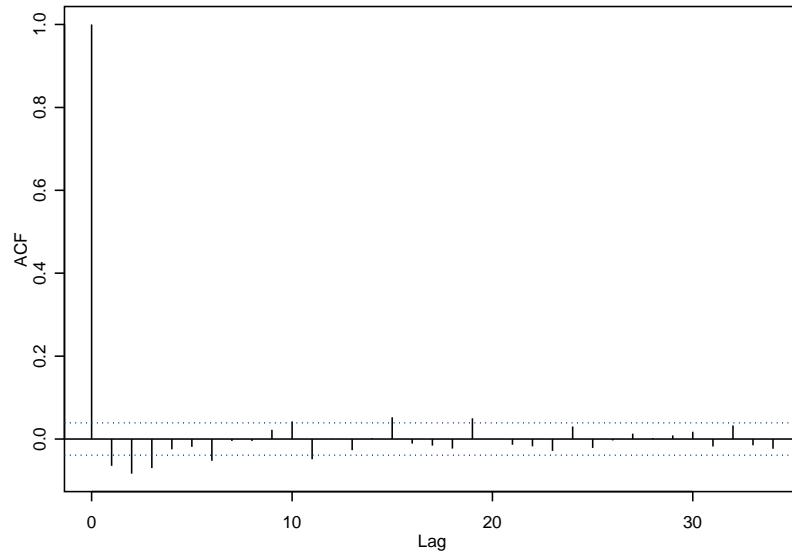


Figure 2.3: Autocorrélations des rendements arithmétiques et géométriques de GM.

Series : Mobile\$Rendement.g.om.triue



Series : Mobile\$squared

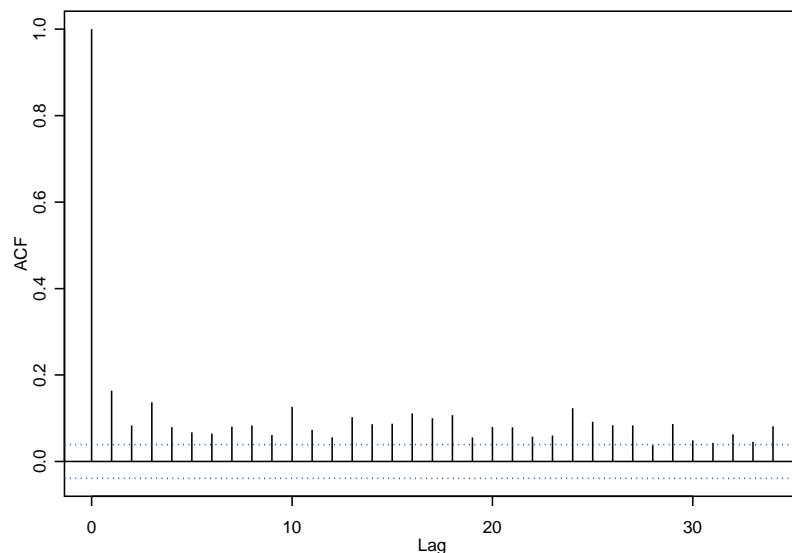
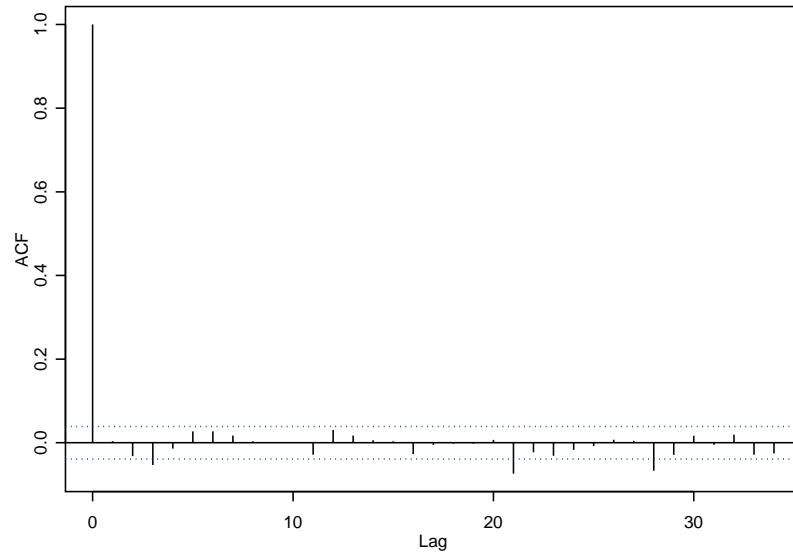


Figure 2.4: Autocorrélations des rendements arithmétiques et géométriques de Mobil.

Series : Microsoft\$Rendement.g.om.triue



Series : Microsoft\$Squared

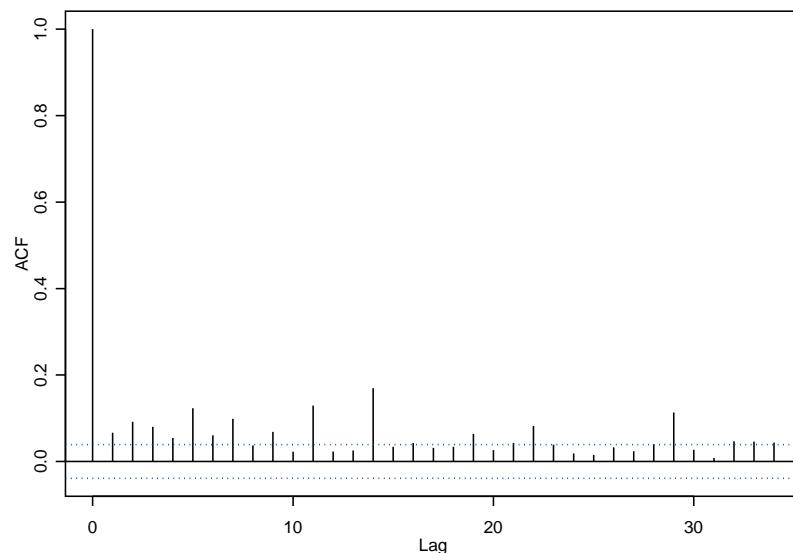


Figure 2.5: Autocorrélations des rendements arithmétiques et géométriques de Microsoft.

où $\hat{\mu} = \frac{1}{N} \sum_{t=1}^T u_t$

Dans le cas des carrés des rendements, u_t est remplacé dans la formule précédente par u_t^2 et donc $\hat{\rho}(k)$ devient:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^T (u_t^2 - \hat{\sigma})(u_{t-k}^2 - \hat{\sigma}^2)}{\sum_{t=1}^T (u_t^2 - \hat{\sigma})^2}$$

où $\hat{\sigma}$ est le moment d'échantillonnage d'ordre 2:

$$\hat{\sigma} = \frac{1}{N} \sum_{t=1}^T u_t^2$$

Dans l'hypothèse où il n'y aurait pas d'autocorrélations, le test statistique suit asymptotiquement une loi $\chi^2(q)$.

2.2.5 La volatilité

Il n'est pas évident de définir ce qu'est la volatilité d'une série de rendement. Traditionnellement, la déviation standard des rendements a été utilisé comme mesure de la volatilité. Cette mesure est facile à estimer et à interpréter tant qu'elle reste constante. Malheureusement, comme nous l'avons démontré auparavant, la déviation standard des rendements d'un actif financier n'est pas constante. Le niveau de volatilité d'aujourd'hui peut être différent de celui d'hier, ce qui nous pousse à penser à une manière de mesurer cette volatilité instantanée.

2.2.6 Conclusion du chapitre

Dans ce chapitre nous avons présenté quelques caractéristiques des séries de rendements. Parmi ces caractéristiques, on retrouve l'excès du coefficient d'aplatissement, le coefficient d'asymétrie (skewness) ainsi que l'hétéroscédasticité et l'autocorrélation dans le carré des rendements. Nous avons montré les techniques, soit visuelles ou statistiques, utilisées pour vérifier la présence de ces caractéristiques. Dans le prochain chapitre nous introduirons les modèles utilisés pour modéliser les actifs financiers, en particulier le modèle de Black-Scholes qui reste le plus utilisé dans le domaine de l'ingénierie financière, ainsi que les différentes approches utilisées pour estimer la volatilité.

Chapitre 3

Modèle de Black-Scholes et estimation de la volatilité

3.1 Présentation du modèle de Black-Scholes

On se place dans un marché financier composé d'un actif sans risque dont le prix ($S_t^0 = e^{rt}$)¹ obéit à l'équation différentielle:

$$\begin{cases} S_0^0 = 1 \\ dS_t^0 = rS_t^0 dt \end{cases} \quad (3.1)$$

et d'un actif avec risque dont le prix S_t^1 vérifie l'équation différentielle stochastique:

¹En général, r correspond au taux d'un bond du trésor de même échéance que l'option.

$$\begin{cases} S_0^1 > 0; \\ dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 dB_t \end{cases} \quad (3.2)$$

tel que μ et σ sont deux constantes; σ non nulle et B_t un mouvement brownien réel ². L'équation(3.2) se résout explicitement grâce à la formule d'Itô³, en considérant la variable $Y_t = \log(S_t)$. La solution est donnée par l'expression suivante:

$$S_t^1 = S_0^1 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t \right).$$

On désigne par $(\Omega, \mathcal{F}_t, P)$ l'espace probabilisé filtré sur lequel est construit B_t . On définit une option européenne f sur l'actif (S_t) , d'échéance T , comme étant une variable aléatoire sur Ω qui est \mathcal{F}_T -mesurable et positive. Par exemple, les options européennes où $f(\omega) = f(S_T(\omega))$, c'est à dire que f ne dépend que de la valeur finale de l'actif sous-jacent. On suppose que le marché financier est complet et qu'il n'y a pas de possibilité d'arbitrage.

Exemple: $f = (S_T - K)^+$ pour une option d'achat “call” européen et $f = (K - S_T)^+$ pour une option de vente “put” européen. La valeur V_t au temps $t \in [0, T]$ d'une option européenne f sur l'actif S_t , d'échéance T , est donnée par:

$$V_t = E^*(e^{-r(T-t)} f | \mathcal{F}_t),$$

où E^* représente l'espérance au sens d'une probabilité, dite probabilité de calcul, pour laquelle l'actif sous-jacent actualisé $\tilde{S}_t = e^{-rT} S_t$ est une martin-

²Voir Annexe1 pour la définition d'un mouvement brownien

³Annexe1

gale. Sous cette probabilité, nous avons:

$$dS_t^1 = rS_t^1 dt + \sigma S_t^1 dB_t^* \quad (3.3)$$

Dans le modèle de Black-Scholes, où on considère que le prix de l'option, une option d'achat d'échéance T et de prix d'exercice K par exemple, est une fonction régulière de la maturité τ et de la valeur y de l'actif sous-jacent, on montre que ce prix $C(\tau, y)$ est justement l'espérance des valeurs actualisées que peut prendre l'option à l'échéance, c'est à dire les valeurs yS_τ ; donc

$$\begin{aligned} C(\tau, y) &= E(f(y\tilde{S}_\tau)) = E(f(e^{-r\tau} yS_\tau)) \\ &= e^{-r\tau} E((yS_\tau - K)^+) \end{aligned}$$

avec $f(S) = (S - K)^+$ dans le cas d'un call. Et justement, on montre que $C(\tau, y)$ est solution de l'équation, dite équation aux dérivées partielles de Black-Scholes:

$$\begin{cases} \frac{\partial C}{\partial \tau} = \frac{1}{2}\sigma^2 y^2 \frac{\partial^2 C}{\partial \tau^2} + ry \frac{\partial C}{\partial y} - rC; \\ C(0, y) = (y - K^+) \end{cases}$$

dont la solution explicite donne la formule de Black-Scholes :

$$C(\tau, S) = S\mathcal{N}(d_1) - Ke^{-r\tau}\mathcal{N}(d_2)$$

où

$$\begin{cases} d_1 = \frac{1}{\sigma\sqrt{\tau}}(\log \frac{S}{K} + (r + \frac{\sigma^2}{2})\tau); \\ d_2 = \frac{1}{\sigma\sqrt{\tau}}(\log \frac{S}{K} + (r - \frac{\sigma^2}{2})\tau); \\ \mathcal{N}(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{\xi^2}{2}} d\xi \end{cases}$$

3.2 Importance de la volatilité

Lorsqu'un produit financier suit la loi

$$dS_t = S_t(\mu dt + \sigma dB_t),$$

on a coutume de dire que σ représente la volatilité de ce produit. Intuitivement σ représente l'écart-type de $\frac{dS}{S}$ et est lié au risque de l'actif. Il est à remarquer que le coefficient μ (*appreciation rate*) n'entre pas dans la formule d'évaluation de Black-Scholes. On peut donc prendre μ qui varie dans le temps ou même qui soit un processus stochastique adapté à une certaine filtration.

La formule de Black&Scholes ne donne un résultat explicite que lorsque les coefficients r et σ sont déterministes. Lorsque la volatilité est aléatoire, cela implique une incomplétude du marché et le prix de l'option est donné à ce moment là par une espérance conditionnelle, c'est ce qu'on verra par la suite de ce mémoire.

La formule d'Itô appliquée à $C(t, S_t)$ conduit à

$$\begin{aligned} dC_t &= \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial x} dS_t + \frac{1}{2}(\sigma S_t)^2 \frac{\partial^2 C}{\partial x^2} dt \\ &= \left(\frac{\partial C}{\partial t} + \frac{\partial C}{\partial x} \mu S_t + \frac{1}{2}(\sigma S_t)^2 \frac{\partial^2 C}{\partial x^2} \right) dt + \frac{\partial C}{\partial x} \sigma S_t dB_t \end{aligned} \tag{3.4}$$

et la volatilité de l'option d'achat est $\frac{1}{C} \frac{\partial C}{\partial x} S \sigma$.

Ce que nous pouvons écrire

$$\nu_c = \frac{S}{C} \frac{\partial C}{\partial x} \nu_S$$

En posant $\eta = \frac{S}{C} \Delta$, on obtient que la volatilité de l'option d'achat est proportionnelle à la volatilité de l'actif sous-jacent: $\nu_c = \eta \nu_S$. De plus $\nu_c \geq \nu_S$ ce qui justifie l'achat d'une option d'achat. Quand on compare la valeur d'une option calculée en utilisant le modèle de Black-Scholes avec le prix réel de l'option, il existe généralement une différence entre les deux valeurs. Ce biais dans le modèle d'évaluation est connu comme étant l'effet “Smile”: La formule d'évaluation d'option de Black-Scholes tend à sous-estimer les options hors jeu (“out-of-the-money”), et à sur-estimer les options en jeu (“at-the-money”), ce qui implique que la volatilité implicite change avec le prix d'exercice (“Strike price”). Ceci est dû à la supposition faite dans le modèle de Black-Scholes comme quoi la volatilité est constante. En réalité, la volatilité n'est pas constante. Elle n'est pas non plus uniforme. Toutes ces observations nous amènent à considérer la volatilité comme variable aléatoire. La question est de savoir comment modéliser la volatilité stochastique.

3.3 La notion d'option

3.3.1 Définition

Une option est un titre qui donne à son détenteur le droit, mais non l'obligation, d'acheter (ou de vendre) un titre à un prix déterminé à l'avance (le prix

d'exercice ou prix de levée de l'option) pendant une certaine période de temps. Passé cette période, l'option ne vaut plus rien.

3.3.2 Les deux formes d'options

Les options d'achat(“Call”)

Une option d'achat donne à son détenteur le droit, mais non l'obligation, d'acheter un titre à un prix déterminé à l'avance, appelé prix d'exercice, pendant une période de temps déterminée. Une fois passée sa date d'échéance, l'option d'achat ne vaut plus rien. L'option d'achat est donc un bien périssable.

Les options de vente(“Put”):

Une option de vente, quant à elle, donne à son détenteur le droit, mais non l'obligation, de vendre un titre à un prix déterminé à l'avance, pendant une période de temps déterminée.

3.3.3 Option européenne et option américaine

Si l'option peut être exercée à n'importe quel instant précédent l'échéance, on parle d'option américaine, si l'option ne peut être exercée qu'à l'échéance, on parle d'option européenne.

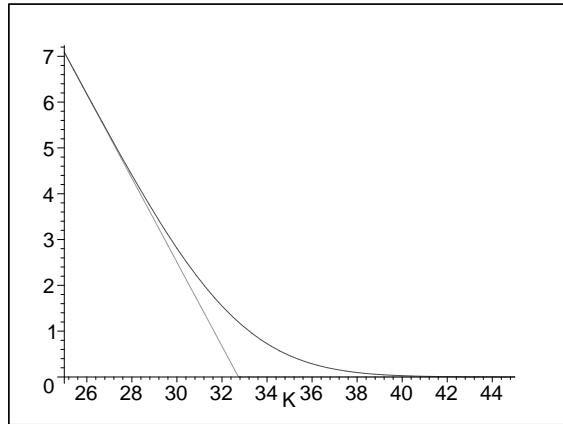


Figure 3.1: Évolution de la valeur d'une option de vente en fonction du prix d'exercice

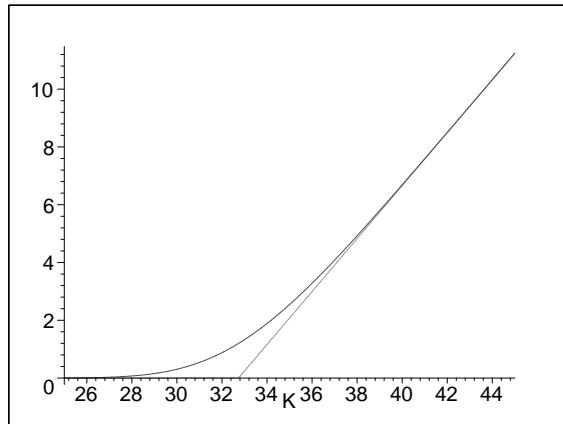


Figure 3.2: Évolution de la valeur d'une option d'achat en fonction du prix d'exercice

3.3.4 La notion d'arbitrage et la relation de parité call-put

L'hypothèse de base, retenue dans tous les modèles, est que, dans un marché suffisamment actif, il n'y a pas d'opportunité d'arbitrage, c'est à dire qu'il est impossible de faire des profits sans prendre de risques de manière générale.

Exemple élémentaire d'arbitrage

Supposons qu'un titre vaut 15\$ dans un an. Aujourd'hui, ce titre vaut 10\$, et le taux d'intérêt pour l'emprunt d'une banque vaut 4%. On peut emprunter 1000\$ de la banque, et acheter 100 actions. Dans un an, on revend le titre à 15 \$ l'action et on rembourse l'emprunt, ce qui nous fait un bénéfice de $1500-1040=460$ \$ sans prendre de risque.

Désignons par C_t et P_t les prix respectifs du call et du put à l'instant t . En l'absence d'opportunité d'arbitrage, on a la relation suivante, valable à tout instant $t < T$ et appelée “relation de parité call-put”:

$$C_t - P_t = S_t - Ke^{-r(T-t)},$$

où K est le prix d'exercice et S_t est le cours d'une action à la date t .

Telle que définit auparavant, la volatilité d'une action, est la mesure de notre incertitude sur les gains sur cette action. Les valeurs typiques de la volatilité d'une action sont entre 20% et 40% par an.

3.4 Estimation de la volatilité à partir de données historiques

Pour estimer empiriquement la volatilité du prix d'une action, on doit observer le prix de l'action en question dans des intervalles de temps fixe(ex: chaque jour, chaque semaine ou chaque mois).

On définit alors:

n : Nombre d'observations;

S_i : Le prix de l'action à la fin du i^{eme} intervalle

τ : Longueur de l'intervalle de temps en années

et soit:

$$u_i = \log\left(\frac{S_i}{S_{i-1}}\right) \text{ pour } i = 1, \dots, n.$$

L'estimateur usuel, s , de la déviation standard des u_i est donné par:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2}$$

ou

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n u_i^2 - \frac{1}{n(n-1)} \left(\sum_{i=1}^n u_i\right)^2},$$

où \bar{u} est la moyenne des u_i .

Or d'après ce qui précède, la déviation standard des u_i est $\sigma\sqrt{\tau}$. Donc la

variable s estime $\sigma\sqrt{\tau}$. d'où on peut estimer σ par σ^* tel que:

$$\sigma^* = \frac{s}{\sqrt{\tau}}$$

On peut facilement montrer que l'erreur standard de cette estimation est approximativement $\frac{\sigma^*}{\sqrt{2n}}$.

Le choix de la valeur n n'est cependant pas facile, mais plus le nombre de données est grand, meilleure sera l'approximation. Cependant, on sait maintenant que σ change au cours du temps et de vieilles données peuvent ne pas être révélatrice de la tendance future de la volatilité. Dans Hull (1999), il est conseillé de prendre les données entre 90 et 180 jours. Cependant, si on cherche à estimer la volatilité afin de calculer le prix d'exercice d'une option sur deux ans, il faut prendre en considérations les données des deux dernières années. Comme nous le verrons dans un prochain chapitre, d'autres méthodes plus sophistiquées sont utilisées. Les prochains tableaux illustrent les résultats de ces calculs pour les 4 titres considérés.

Des recherches empiriques ont montré qu'il faut utiliser le nombre de jours ouvrables comme mesure de temps dans l'estimation.

Nombre de jours (n)	Volatilité historique(%)	Erreur standard
40	22.6	0.025
50	21.4	0.021
60	20.8	0.019
70	21.5	0.018
80	24.5	0.019
90	24.2	0.018
100	25.6	0.018
110	26.4	0.017
120	26.4	0.017
1000	26.2	0.005
1500	27.5	0.005
2000	31.2	0.004
2527	36.0	0.005

Tableau 3.1: Volatilité historique et erreur standard du titre de Coca-Cola Company en fonction de l'historique (n).

Nombre de jours (n)	Volatilité historique(%)	Erreur standard
40	21.5	0.024
50	30.0	0.030
60	30.0	0.027
70	30.8	0.026
80	36.5	0.028
90	37.7	0.028
100	37.7	0.026
110	38.4	0.025
120	38.0	0.026
1000	30.4	0.006
1500	28.5	0.005
2000	29.7	0.004
2527	31.4	0.004

Tableau 3.2: Volatilité historique et erreur standard du titre de General Motors Company en fonction de l'historique (n).

Nombre de jours (n)	Volatilité historique(%)	Erreur standard
40	20.3	0.022
50	18.6	0.018
60	17.5	0.016
70	17.5	0.014
80	17.4	0.013
90	17.8	0.013
100	18.1	0.012
110	19.4	0.013
120	20.9	0.013
1000	27.1	0.006
1500	25.6	0.004
2000	23.3	0.003
2527	22.2	0.003

Tableau 3.3: Volatilité historique et erreur standard du titre de Mobil Oil Company en fonction de l'historique (n).

Nombre de jours (n)	Volatilité historique(%)	Erreur standard
40	36.1	0.040
50	39.9	0.039
60	38.15	0.034
70	36.3	0.030
80	35.1	0.027
90	34.0	0.025
100	35.8	0.025
110	38.8	0.026
120	41.3	0.026
1000	44.3	0.009
1500	40.8	0.007
2000	38.3	0.006
2527	37.1	0.005

Tableau 3.4: Volatilité historique et erreur standard du titre de Microsoft en fonction de l'historique (n).

3.5 La volatilité implicite

3.5.1 Définition

Selon De la Bruslerie (1988): La volatilité implicite est le résultat d'une équation liant le prix au marché d'une option à ses déterminants dans le cadre d'un modèle d'évaluation donné . Elle peut s'interpréter comme étant

une estimation contemporaine de la variabilité moyenne future de l'actif sous-jacent par le marché au cours de la vie de l'option. La volatilité implicite serait donc une estimation actuelle ayant un contenu prospectif, toujours selon De la Bruslerie (1988).

3.5.2 Modèle d'évaluation

Traditionnellement, le modèle de Black-Scholes (1973) a été utilisé pour estimer la volatilité implicite d'une option donnée. Comme nous l'avons déjà vu, les auteurs de ce modèle supposent que cette variable est constante durant la vie de l'option, ce qui n'est vrai dans la réalité. L'utilisation d'un modèle d'évaluation d'options pour estimer la volatilité implicite suppose le modèle connu et bien spécifié, c'est à dire que le prix que dégage le modèle est le prix de l'option sur le marché. Donc, la mesure de la volatilité implicite est extrêmement tributaire du choix du modèle d'évaluation d'options que nous choisissons pour l'estimer. Afin de minimiser l'erreur de l'estimation, il faudrait donc choisir un modèle qui reflète la réalité de notre actif sous-jacent.

3.5.3 Calcul de la volatilité implicite

Dans le modèle de Black-Scholes qu'on a vu à la section 1 du présent chapitre, la valeur d'une option call européenne est donnée par:

$$C(\tau, S) = S\mathcal{N}(d_1) - Ke^{-r\tau}\mathcal{N}(d_2) \quad (3.5)$$

où

$$\begin{cases} d_1 = \frac{1}{\sigma\sqrt{\tau}}(\log \frac{S}{K} + (r + \frac{\sigma^2}{2})\tau); \\ d_2 = \frac{1}{\sigma\sqrt{\tau}}(\log \frac{S}{K} + (r - \frac{\sigma^2}{2})\tau); \\ \mathcal{N}(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{\xi^2}{2}} d\xi \end{cases}$$

La valeur C est donnée en fonction de six paramètres, dont σ est le seul paramètre qui ne soit pas directement observable. Si on connaît σ on peut calculer le prix de l'option. Inversement si on connaît la valeur de l'option, on peut déduire la valeur de σ , car $\frac{\partial C}{\partial \sigma} > 0$. Étant donnée que la valeur de l'option à l'instant t est comprise entre $\max(S - Ke^{-r\tau}, 0)$ et S , alors il existe une et une seule valeur de σ telle que la valeur théorique de l'option soit égale à la valeur réel sur le marché de cette même option. La valeur trouvée, est ce qu'on appelle la volatilité implicite.

À titre d'exemple, on suppose que la valeur d'un call sur un titre qui ne paye pas de dividendes soit 1.875 tel que

$$S_0 = 21, X = 20, r = 0.1 \text{ et } T = 0.25.$$

La volatilité implicite est la valeur de σ , telle que si on la remplace dans l'équation (3.5), on retrouve la valeur $C = 1.875$. Malheureusement on ne peut inverser l'équation (3.5) pour avoir une expression de σ en fonction de S_0 , X , r , τ et C . Cependant, un procédé de recherche itératif, tel que l'algorithme de Raphson, peut être utilisé pour retrouver la valeur de la volatilité implicite σ . Par exemple, on commence en essayant la valeur $\sigma = 0.20$. Ce qui donne à C la valeur 1.76, ce qui est en dessous de la valeur visée de 1.875. Comme C est une fonction croissante par rapport à σ , il nous faut prendre une valeur plus grande pour σ . On peut essayer comme deuxième

valeur 0.30 ce qui donne à C la valeur 2.10, ce qui est au-dessus de la valeur recherchée. Donc la valeur recherchée se trouve entre 0.20 et 0.30. Par un procédé itératif, on divise l'intervalle $[0.20, 0.30]$ en deux jusqu'à arriver à la valeur recherchée. Dans cet exemple, la volatilité implicite est 0.235 ou 23.5% par an.

Tel que montré par plusieurs auteurs, Latané et Rendleman (1976), Beckers (1981), l'estimation de la volatilité actuelle basée sur la volatilité implicite est plus performante, en terme de prédiction, par rapport aux estimations basées sur la volatilité historique.

3.5.4 Facteur influençant la volatilité implicite

Rubinstein (1985) a examiné les transactions sur les options d'achat à partir du “Berkeley Options Data Base” pour conduire des tests non paramétriques de l'hypothèse nulle de Black-Scholes, qui stipule que la volatilité implicite ne montre pas de différences systématiques selon le prix d'exercice ou selon la maturité pour des options identiques. Les résultats les plus pertinents de Rubenstein démontrent que la volatilité implicite des options d'achat hors jeu est systématiquement plus élevée pour les options proches de l'échéance.

Morse (1991) a étudié l'effet des jours de la semaine sur les volatilités implicites des options sur indice boursier et ce en isolant le comportement quotidien des volatilités implicites. L'auteur a trouvé que pour les options sur indice la différence entre la volatilité implicite des options d'achat et celle des options de vente, tend à baisser le vendredi et à augmenter le lundi. Pour les

options sur actions individuelles, des effets similaires ne sont pas évidents. Aussi bien pour les indices que pour les titres individuels, la volatilité implicite diminue le mardi, le mercredi et le jeudi. La variance tend à être plus élevée le vendredi et le lundi.

Derman, Kani et Zou (1996) ont présenté le concept de “Local Volatility” ou volatilité locale et ses utilisations. Les volatilités locales sont utilisées dans les marchés avec une structure de volatilité “Smile” prononcée pour mesurer le sentiment du marché, pour calculer l’évolution des volatilités implicites dans le temps.

Xu et Taylor (1994a) ont examiné la structure à terme de la volatilité des options sur devises. Ils ont décrit différentes méthodes pour l'estimation de cette structure et ont proposé un modèle de la structure à terme de la volatilité anticipée ainsi que des caractéristiques des séries chronologiques de la structure à terme. Les auteurs ont utilisé des données relatives aux prix de clôtures des options sur devises du “Philadelphia Stock Exchange”, de la période allant de 1985 jusqu'à 1989. Les résultats ont été obtenus pour les options d'achat et de vente de type américain sur quatre devises différentes: la livre sterling, le deutsche mark, le yen japonais et le franc suisse. Xu et Taylor ont abouti à cinq conclusions principales: La présence d'effets significatifs de la structure à terme qui font que la volatilité implicite varie fortement selon les maturités, la pente de la structure à terme varie parfois vers le haut parfois vers le bas et sa direction change fréquemment, la présence de variations significatives dans les anticipations de la volatilité à long terme, les structures à terme à chaque instant des 4 devises sont

très similaires, et concluent par l'absence de stationnarité des éléments de la structure à terme.

3.5.5 Techniques d'estimation de la volatilité implicite

Il existe plusieurs façons d'estimer la volatilité implicite: selon le modèle d'évaluation, les types, les classes et les séries d'options disponibles. Cela pose un problème car intuitivement, il devrait exister une seule mesure de la variabilité par actif financier sous-jacent à l'option. Pour résoudre le problème de la structure de la volatilité implicite, plusieurs auteurs ont suggéré différentes méthodes de calcul afin “d'uniformiser” ou de trouver un seul chiffre pouvant caractériser un actif financier en particulier. Dans sa revue de littérature, Mayhew (1995) expose trois façons d'estimer cette mesure de la variabilité unique. Toutes ces techniques ont ceci en commun: elles utilisent une moyenne pondérée de la volatilité implicite des options se transigeant sur un même actif financier.

La première, qui a été suggérée par Trippi (1977) et par la suite par Schmalensee et Trippi (1978), calcule la volatilité implicite (σ_{impl}) pour toutes les options d'une même classe et de là, calcule une simple moyenne arithmétique de σ_{impl} comme ceci:

$$\hat{\sigma}_{impl} = \frac{1}{N} \sum_{i=1}^N \sigma_{impl}$$

L'idée derrière ce raisonnement consiste à dire que si le modèle évalue correctement le prix d'une option sur le marché, alors les déviations de la prévision du modèle par rapport au marché s'expliquent par une forme de

bruit dans le marché, une distorsion qu'il est possible de réduire en augmentant le nombre d'observations dans le calcul du σ_{impl} moyen. Or, nous savons qu'un modèle comme celui de Black-Scholes a tendance à surévaluer les options à parité et à sous-évaluer celles qui sont fortement hors-jeu ou en jeu, Khoury et Laroche (1995). De plus, certaines options sont plus sensibles à la volatilité que d'autres en raison de différents facteurs comme l'asynchronisme dans les prix, ce qui baisse encore plus l'évaluation de la prime.

Latané et Rendleman (1976) suggèrent l'équation suivante pour tenir compte de ce phénomène:

$$\hat{\sigma}_{impl} = \frac{1}{\sum_{i=1}^N w_i} \sqrt{\sum_{i=1}^N w_i^2 \sigma_{impl}^2}$$

où w_i est le vega de l'option calculé avec la formule de Black-Scholes. Le coefficient vega mesure la sensibilité du prix d'une option à la variation de la volatilité du titre sous-jacent et se calcule comme suit:

$$\vartheta = P_s \times \sqrt{\tau} \times \frac{e^{-d_1^2/2}}{\sqrt{2\tau}},$$

où

$$\begin{cases} P_s : \text{le prix de l'actif sous-jacent;} \\ \tau : \text{la maturité c;} \\ d_1 = \frac{\ln(\frac{P'}{X}) + (r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}} \end{cases}$$

Cette technique possède l'avantage de pondérer les options selon leur degré de sensibilité à la volatilité. Cependant, elle n'alloue malheureusement pas une somme des poids égale à un. Pour régler ce problème, Manaster (1978) suggère de pondérer σ_{impl} par l'élasticité de la volatilité, ce qui donnera une pondération totale égale à un.

Beckers (1981) et Whaley (1982) ont utilisé une autre forme pondérée afin de réduire la distorsion causée par l'écart entre le prix prévu par le modèle et le prix réel de l'option:

$$\hat{\sigma}_{impl} = \sum_{i=1}^N w_i [C_i - BS_i(\hat{\sigma})]^2,$$

où C_i est la prime au marché de l'option et BS est la prime évaluée par le modèle de Black-Scholes.

Les études récentes sur le sujet de l'estimation de la volatilité implicite s'inspirent de ces quatre méthodes présentées. Il existe donc un éventail de choix possibles de méthodes de mesure de la volatilité implicite. Il faudrait choisir la technique appropriée aux caractéristiques de l'actif financier de cette étude, choisir celle qui améliorera la performance prévisionnelle de la stratégie d'investissement qui sera retenu.

Chapitre 4

Modélisation ARCH

4.1 Introduction aux modèles ARCH

La classe des modèles conditionnellement hétérosclélastique, “*autoregressive conditional heteroscedastic ou ARCH*”, a été proposée par Engle en 1982. La classe de modèles ARCH a été utilisée pour prévoir la volatilité à cause de leur capacité à modéliser des séries financières. Plusieurs auteurs, dont Akgiray (1989), ont montré que les séries temporelles de rendement présentent de l'autocorrélation, que leur distribution est souvent leptokurtique et asymétrique, Bollerslev et Engle (1993), et surtout qu'elles possèdent une variance non constante.

Le recours à cette méthode, s'est fait aussi à cause de l'incapacité des modèles linéaires de type ARMA, utilisés auparavant, de prendre en considération l'information passée contenue dans la série chronologique. D'un point de vue

statistique, les modèles ARCH constituent une classe spécifique de modèles non linéaires pour laquelle on peut mener une étude complète en abordant un certain nombre de problèmes classiques: test de marche aléatoire, détermination d'intervalles de prévision, recherche de séries sous-jacentes, ...

Par définition, une variance conditionnelle au temps t dépend de ce qui s'est passé au temps $t - 1, t - 2, \dots$; tandis qu'une variance inconditionnelle est fixe et n'est pas reliée au passé. Cette caractéristique de la variance a été également vérifiée empiriquement. Christie (1982) a montré en effet que l'arrivée d'information nouvelle sur le marché influence la volatilité d'un actif financier. On remarque aussi que la volatilité d'un titre boursier à tendance à augmenter à l'approche de l'annonce des résultats financiers de l'entreprise. Christie (1982) remarque ainsi que les taux de rendement sur actions sont négativement corrélés avec les changements que subit la variance.

4.2 Présentation théorique des modèles de types ARCH

Cette section est dévolue à la présentation et à la discussion des modèles de type ARCH. Nous nous limitons ici à présenter les classes les plus utilisées de ces modèles, mais il faut toutefois être conscient du volume croissant des modèles développés et des applications de ces modèles réalisés dans les domaines de la finance et de l'économie. Pour une revue complète et exhaustive de l'ensemble de ces modèles, nous renvoyons le lecteur aux travaux de Boller-

slev, Chou et Kroner (1992), Palm (1996) et ceux de et à ceux de Ghysels, Harvey et Renault (1996) et Shepard (1996) pour une revue plus générale sur les modèles à volatilité stochastique.

4.2.1 Processus ARCH

Le premier modèle de type a été initialement utilisé par Engle (1982) pour estimer la variance de l'inflation en Grande-Bretagne. L'aspect hétéroscédastique est introduit par l'intermédiaire d'une dynamique autorégressive sur les carrés du bruit, Engle (1982). Le modèle est défini par:

$$\begin{cases} Y_t = \varepsilon_t, \\ \text{avec : } E(\varepsilon_t | \mathcal{F}_{t-1}) = 0 \\ E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = h_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2, a_i \geq 0, \end{cases}$$

où q est l'ordre du processus ARCH et $\mathcal{F}_{t-1} = \sigma\{\varepsilon_1, \dots, \varepsilon_{t-1}\}$, qui représente le nombre de retards retenu au niveau du terme d'erreur.

La volatilité de la date t est donc fonction des carrés des écarts à la moyenne observés dans le passé proche. Si les coefficients $a_i, i = 1, \dots, q$, sont tous positifs et assez grands, il y a une certaine persistance des niveaux volatilité: on observe alors des périodes de fortes volatilités suivies de périodes de faibles volatilités. Si le polynôme caractéristique $1 - \sum_{i=1}^q a_i z^i$ admet des racines complexes, on peut observer des phénomènes cycliques sur les niveaux de volatilité.

Dans la plupart des applications, ε_t représente l'innovation par rapport à la moyenne $g(\Psi_{t-1})$ d'un autre processus stochastique y_t , Ψ_{t-1} étant l'ensemble

informationnel à l'instant $t - 1$:

$$y_t = g(\Psi_{t-1}) + \epsilon_t$$

Prenons, pour illustrer ceci, le modèle autoregressif d'ordre un suivant¹:

$$y_t = \mu + \varphi y_{t-1} + \epsilon_t, \forall t, |\varphi| < 1,$$

avec ϵ_t un bruit blanc (au sens faible) qui satisfait la condition de différence de martingales:

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0, \forall t,$$

et dont la variance conditionnelle, $V(\epsilon_t | \mathcal{F}_{t-1})$, suit un processus ARCH(1) :

$$\epsilon_t^2 = c + \alpha \epsilon_{t-1}^2 + u_t, \quad (4.1)$$

u_t étant un bruit blanc gaussien.

Les propriétés du processus d'innovation ϵ_t sont les suivantes:

1. Le processus ϵ_t est orthogonal aux valeurs passées, quel que soit le retard:

$$E(\epsilon_t | \mathcal{F}_{t-\tau}) = 0, \forall \tau > 0$$

2. La propriété d'orthogonalité implique que les corrélations conditionnelles sont nulles:

$$\begin{aligned} \text{cov}[(\epsilon_t, \epsilon_{t+k}) | \mathcal{F}_{t-\tau}] &= E[(\epsilon_t \epsilon_{t+k}) | \mathcal{F}_{t-\tau}] - E[\epsilon_t | \mathcal{F}_{t-\tau}] E[\epsilon_{t+k} | \mathcal{F}_{t-\tau}] \\ &= E[(\epsilon_t \epsilon_{t+k}) | \mathcal{F}_{t-\tau}] \\ &= E[(\epsilon_t \epsilon_{t+k}) | \mathcal{F}_{t+k-1} | \mathcal{F}_{t-\tau}] \\ &= E[\epsilon_t E(\epsilon_{t+k} | \mathcal{F}_{t+k-1}) | \mathcal{F}_{t-\tau}] \\ &= 0 \end{aligned}$$

¹Cet exemple est tiré de Gourieroux et Monfort ((1995),p.35-36)

où τ et k sont des entiers positifs. Il y a donc absence de corrélation entre les valeurs présentes et futures du processus, quels que soient les retards τ et k ;

3. La variance conditionnelle du processus ϵ_t varie à travers le temps, bien que sa variance non conditionnelle soit supposée constante.

Pour montrer cette dernière propriété, considérons la substitution récursive réalisée dans l'équation (4.1):

$$\epsilon_t^2 = c[1 + \alpha + \dots + \alpha^{\tau-1}] + \alpha^\tau \epsilon_{t-\tau}^2 + u_t + \alpha u_{t-1} + \dots + \alpha^{\tau-1} u_{t-\tau+1}$$

L'espérance de ϵ_t^2 conditionnelle à l'ensemble d'information $\epsilon_{t-\tau}$ est alors:

$$E[(\epsilon_t^2 \epsilon_{t-\tau})] = c[1 + \alpha + \dots + \alpha^{\tau-1}] + \alpha^\tau \epsilon_{t-\tau}^2 = c \left(\frac{1 - \alpha^\tau}{1 - \alpha} \right) + \alpha^\tau \epsilon_{t-\tau}^2$$

La variance conditionnelle est donc égale à

$$V(\epsilon_t | \mathcal{F}_{t-\tau}) = h_t = c \frac{1 - \alpha^\tau}{1 - \alpha} + \alpha^\tau \epsilon_{t-\tau}^2$$

et dépend de la valeur la plus récente sur l'information passée, i.e $\epsilon_{t-\tau}^2$.

Lorsque le retard τ tend vers l'infini, la variance conditionnelle converge vers la variance non conditionnelle qui est indépendante du temps:

$$\sigma^2 = V(\epsilon_t) = E[V(\epsilon_t | \mathcal{F}_{t-\tau})] = \frac{c}{1 - \alpha}$$

Le modèle ARCH(1) peut être généralisé de manière à inclure des paramètres autorégressifs additionnels.

Définition. Le modèle $ARCH(Q)$ proposé par Engle(1982) est défini comme:

$$\epsilon_t^2 = c + \sum_{i=1}^Q \alpha_i \epsilon_{t-i}^2 + u_t, \quad E(\epsilon_t | \mathcal{F}_{t-1}) = 0,$$

où u_t est une différence de martingales. La variance conditionnelle de ϵ_t est

$$V(\epsilon_t | \epsilon_{t-1}) = h_t = c + \sum_{i=1}^Q \alpha_i \epsilon_{t-i}^2 \quad (4.2)$$

et dépend des Q valeurs les plus récentes de ϵ_t^2 sur l'information passée.

La persistance de la volatilité est ainsi modélisée en spécifiant la variance conditionnelle comme une fonction du carré des innovations passées. Il est nécessaire d'imposer des conditions sur les paramètres α_i dans l'équation (4.2) afin de maintenir une variance conditionnelle positive. En particulier, il faut que tous les paramètres α soient non négatifs:

$$\alpha_i \geq 0, \quad \text{pour } i = 1, \dots, Q$$

Lorsque cette condition est vérifiée, la somme des paramètres autorégressifs du modèle $ARCH(Q)$ doit être inférieur à l'unité pour que le processus ϵ_t^2 soit stationnaire au sens large²:

$$\alpha_1 + \alpha_2 + \dots + \alpha_Q < 1$$

Dans ce cas, la variance non conditionnelle de ϵ_t est donnée par:

$$\sigma^2 = V(\epsilon_t) = \frac{c}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_Q}$$

²Pour que le processus ϵ_t^2 soit stationnaire au sens large, il faut que les racines de $1 - \alpha_1 z - \dots - \alpha_Q z^Q = 0$ soient en dehors du cercle unitaire. Lorsque les α_i sont tous non négatifs, ceci revient à poser la condition que la somme des paramètres autorégressifs du modèle $ARCH$ est inférieur à l'unité.

4.2.2 Processus GARCH

La variance conditionnelle du processus ARCH(Q) décrite dans l'équation (4.2) peut être généralisée à un nombre infini de retards dans le polynôme autorégressif:

$$h_t = c + \pi(L)\epsilon_t^2 \quad (4.3)$$

avec $\pi(L) = \sum_{i=1}^{\infty} \pi_i L^i$. Le polynôme infini $\pi(L)$ peut être formulé comme un rapport de deux polynômes d'ordre fini (Hamilton (1994), p.665):

$$\pi(L) = \frac{\alpha(L)}{1 - \beta(L)} = \frac{\alpha_1 L + \alpha_2 L^2 + \dots + \alpha_Q L^Q}{1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_P L^P}$$

où les racines de $1 - \beta(z) = 0$ sont supposées être en dehors du cercle unitaire.

Si on multiplie l'équation (4.3) par $[1 - \beta(L)]$, on trouve :

$$[1 - \beta(L)]h_t = [1 - \beta(L)]c + \alpha(L)\epsilon_t^2$$

ce qui peut encore s'écrire:

$$h_t = \xi + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \dots + \beta_P h_{t-P} + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_Q \epsilon_{t-Q}^2 \quad (4.4)$$

avec $\xi \equiv [1 - \beta_1 - \beta_2 - \dots - \beta_P]c$

L'équation (4.4) représente le modèle GARCH(P,Q) proposé par Bollerslev (1986), où la variance conditionnelle est une fonction linéaire à la fois du carré des innovations passées et des variances conditionnelles passées. Ce modèle représente une généralisation du processus ARCH permettant de restreindre le nombre de coefficients à estimer et par conséquent le nombre de contraintes.

Définition: Le modèle GARCH(P, Q) proposé par Bollerslev (1986) est défini par:

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0,$$

$$V(\epsilon_t | \mathcal{F}_{t-1}) = h_t = \xi + \sum_{i=1}^Q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^P \beta_i h_{t-i}, \quad (4.5)$$

Le modèle GARCH(P, Q) peut être réécrit sous une forme alternative, à savoir sous la forme d'un modèle ARMA dans la variance conditionnelle. Pour cela, introduisons l'innovation correspondant au carré du processus $u_t = \epsilon_t^2 - h_t$. En remplaçant h_t par $\epsilon_t^2 - u_t$ dans la représentation GARCH décrite dans (4.5), on obtient:

$$\begin{aligned} \epsilon_t^2 - u_t &= \xi + \sum_{i=1}^Q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^P \beta_i (\epsilon_{t-i}^2 - u_{t-i}) \\ \Leftrightarrow \epsilon_t^2 &= \xi + \sum_{i=1}^{\max(P, Q)} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + u_t - \sum_{i=1}^P \beta_i u_{t-i} \end{aligned} \quad (4.6)$$

avec $\alpha_i = 0$ pour $i > Q$ et $\beta_i = 0$ pour $i > P$. L'équation (4.6) représente alors un modèle $ARMA(\max\{P, Q\}, P)$ pour le processus ϵ_t^2 avec un terme d'erreur u_t qui n'a pas nécessairement une variance constante³. Comme pour le processus ARCH(Q), il est nécessaire d'imposer des conditions sur les α_i et les β_i afin d'assurer que la variance conditionnelle soit positive:

$$\xi \geq 0, \alpha_i \geq 0, i = 1, \dots, Q, \beta_i \geq 0, i = 1, \dots, P.$$

³L'intérêt de la transformation du modèle GARCH en un modèle ARMA est que la procédure d'identification du nombre de retards P et Q de Box-Jenkins (1976) peut être utilisée sur le processus ϵ_t^2 (Bollerslev (1988)).

Une fois ces conditions posées, le processus GARCH(P,Q) est stationnaire si la propriété suivante est vérifiée.

Théorème 4.2.1 *Le processus ϵ_t , qui satisfait un modèle GARCH(P, Q) avec des coefficients positifs $\xi \geq 0, \alpha_i \geq 0, i = 1, \dots, Q, \beta_i \geq 0, i = 1, \dots, P$ est asymptotiquement stationnaire au sens large si:*

$$\alpha(1) + \beta(1) = \sum_{i=1}^Q \alpha_i + \sum_{i=1}^P \beta_i < 1. \quad (4.7)$$

4.3 Extensions du modèles GARCH

De nombreuses extensions ont été apportées aux processus GARCH univariés. Pour une description complète de ces extensions, nous renvoyons aux travaux de Bollerslev, Chou et Kroner (1992) et Palm (1996). Nous nous limitons ici à présenter les deux principales extensions, à savoir la modélisation de la persistence infinie de la volatilité conditionnelle par un processus intégré en variance(IGARCH) et les processus d'asymétrie dans la variance(EGARCH). Il existe une autre variante des modèles GARCH très populaire en théorie financière, à savoir le modèle (G)ARCH dans la moyenne ((G)ARCH-M) introduit par Engle, Lilien et Robins (1987) et qui consiste à décrire l'impact de la volatilité conditionnelle sur la moyenne conditionnelle.

A. Processus intégrés en variance

Dans les processus ARMA traditionnels, la condition requise pour la stationnarité est que les racines du polynôme autorégressif soient strictement inférieures à l'unité. Dans le cas où certaines de ces racines ont un module égale à un, le processus est non stationnaire et est appelé processus ARIMA. Le même raisonnement peut se faire pour les processus GARCH. Le processus GARCH(P,Q) est spécifié dans l'équation (4.6) sous la forme d'un modèle ARMA(max(P,Q),P) pour le processus ϵ_t^2 et il est stationnaire au sens large lorsque la somme des paramètres autorégressifs est strictement inférieure à l'unité. Ce processus ARMA pour ϵ_t^2 a une racine unitaire lorsque la condition suivante est vérifiée:

$$\sum_{i=1}^Q \alpha_i + \sum_{i=1}^P \beta_i = 1$$

Le processus satisfaisant cette condition est appelé processus GARCH intégré, noté IGARCH (Engle et Bollerslev (1986)). La définition formelle d'un processus IGARCH est la suivante:

Définition 3.3 *Le processus GARCH(P,Q) caractérisé par les deux premiers moments conditionnels:*

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0$$

$$V(\epsilon_t | \mathcal{F}_{t-1}) = h_t = \xi + \sum_{i=1}^Q \alpha_i \epsilon_{t-i} + \sum_{i=1}^P \beta_i h_{t-i}$$

où $\xi \geq 0, \alpha_i \geq 0, i = 1, \dots, Q, \beta_i \geq 0, i = 1, \dots, P$, et dont le polynôme

$$1 + \sum_{i=1}^Q \alpha_i z^i - \sum_{i=1}^P \beta_i z^i$$

possède $d > 0$ racines unitaires et $\max\{P, Q\} - d$ racines en dehors du cercle unitaires, est:

- intégré en variance d'ordre d si $\xi = 0$,
- intégré en variance d'ordre d avec tendance si $\xi > 0$ (Engle et Bollerslev (1986, p.26)).

Si ϵ_t suit un processus IGARCH, alors la variance non conditionnelle de ϵ_t est infinie, de sorte que ni ϵ_t ni ϵ_t^2 ne satisfont la définition d'un processus stationnaire au sens large.

Considérons, à titre d'exemple, le cas d'un processus (I)GARCH(1,1) décrit dans Nelson (1990). Sous l'hypothèse de normalité conditionnelle, ce processus s'écrit:

$$\epsilon_t | \mathcal{F}_{t-1} \sim N(0, h_t),$$

$$h_t = \xi + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \alpha_1 \geq 0, \beta_1 \geq 0.$$

La prévision de la variance conditionnelle à l'horizon de prévision s a la forme générale suivante:

$$E(h_{t+s} | \mathcal{F}_t) = (\alpha_1 + \beta_1)^s h_t + \xi \left[\sum_{i=0}^{s-1} (\alpha_1 + \beta_1)^i \right]$$

Deux cas peuvent être distingués:

1. lorsque $\alpha_1 + \beta_1 < 1$, le processus (ϵ_t) est stationnaire, ce qui signifie qu'un choc sur la variance conditionnelle h_t a un impact décroissant sur h_{t+s} lorsque s s'accroît et est asymptotiquement négligeable.

2. lorsque $\alpha_1 + \beta_1 = 1$, on a:

$$E(h_{t+s}|\mathcal{F}_t) = h_t + s\xi.$$

Dans ce cas, un choc sur la variance conditionnelle h_t va persister sur tout horizon de prévision.

Ainsi, l'existence d'une racine unitaire dans le processus de volatilité conditionnelle implique une persistence infinie des chocs de volatilité. Ceci signifie que l'accumulation d'informations est très importante pour l'estimation de la variance conditionnelle.

B. Processus d'asymétrie dans la variance

Dans les modèles décrits précédemment, les innovations positives et négatives de même ampleur sont supposées avoir un effet symétrique sur la volatilité conditionnelle. En particulier, ces modèles supposent que l'ampleur et non pas le signe des variations anticipées détermine la volatilité, i.e les innovations au carré ϵ_t^2 affectent la variance conditionnelle h_t , de sorte que h_t est invariant au signe des innovations. Or, diverses études ont montré que la plupart des séries financières sont fortement asymétriques, 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 ampleur (Black (1973), Sentana (1991), Campbell et Hentschel (1992), Glosten et al. (1993)).

De nombreux modèles se sont développés afin de permettre cette asymétrie: le modèle GARCH augmenté(GAARCH) de Bera et Lee (1990), le modèle (G)ARCH exponentiel (EARCH ou EGARCH) de Nelson (1991), le modèle

GARCH quadratique (QGARCH) de Engle (1990) et Sentana (1991), le modèle GARCH asymétrique de Engle et Ng (1993), le modèle à puissance asymétrique (APARCH) de Ding, Engle et Granger (1993), le modèle GARCH absolu (AGARCH) de Hentschel (1994) ou encore le modèle GARCH à seuil (TGARCH) de Zakoian (1994)⁴. Dans ces modèles, la variance conditionnelle h_t répond de manière asymétrique aux innovations positives et négatives.

Le modèle le plus utilisé en pratique est le modèle EGARCH de Nelson (1991) qui se présente comme suit:

$$\log(h_t) = \xi + \sum_{i=1}^Q \alpha_i (\varphi \varepsilon_{t-i} + \phi |\varepsilon_{t-i}| - \phi E|\varepsilon_{-i}|) + \sum_{i=1}^P \beta_i h_{t-i},$$

où les paramètres ξ , α_i et β_i ne sont pas contraints à être non négatifs⁵. Le paramètre φ permet l'asymétrie. Si $\varphi = 0$, une innovation positive ($\varepsilon_{t-i} > 0$) a le même effet sur la volatilité qu'une innovation négative de même ampleur. Si $-1 < \varphi < 0$, une innovation positive accroît moins la volatilité qu'une innovation négative. Enfin si $\varphi < -1$, une innovation positive réduit la volatilité, tandis qu'une innovation négative a tendance à l'augmenter.

⁴Palm (1996) présente une revue complète de l'ensemble de ces processus d'asymétrie dans la variance.

⁵Un des avantages du modèle EGARCH est que la variance conditionnelle h_t est mise sous forme logarithmique. Par conséquent, il n'est pas nécessaire d'imposer des restrictions sur les coefficients pour que celle-ci soit positive.

4.3.1 Effets ARCH et coefficient d'aplatissement

Nous examinons si les modèles de type ARCH parviennent à rendre compte de toute l'aplatissement présent dans les séries financières à haute fréquence.

Une des propriétés des modèles de type ARCH est que le coefficient d'aplatissement a tendance à s'accroître lorsqu'on passe d'une distribution conditionnelle à une distribution non conditionnelle. En particulier, Milhøj (1985), Bollerslev (1986) ainsi que Engle et Bollerslev (1986) ont montré que la distribution non conditionnelle des ϵ_t issus des modèles GARCH ou EGARCH avec des erreurs conditionnelles normales a un coefficient d'aplatissement excessif.

Si on considère un processus GARCH conditionnellement gaussien, le coefficient de d'aplatissement, τ_2 , est égal à :

$$\tau_2 = \frac{E(\epsilon_t^4)}{(E(\epsilon_t^2))^2} = 3 + 3 \frac{V[E(\epsilon_t^2 | \mathcal{F}_{t-1})]}{(E(\epsilon_t^2))^2},$$

et est ainsi directement lié à une mesure de l'hétéroscédasticité conditionnelle⁶.

⁶Nous renvoyons à Gouriéroux (1992, p.46) pour la démonstration

4.4 Les processus hétéroscédastiques comme approximations de processus en temps continu

4.4.1 Convergence des chaînes de Markov vers des diffusions

Discrétisation d'Euler d'une diffusion

Soit sur R^p l'équation différentielle stochastique(EDS) suivante:

$$\begin{cases} dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t, t \leq T \\ Y(0) = y_0 \end{cases}$$

où W_t est un mouvement brownien d-dimensionnel.

Comme dans le cas déterministe, une discrétisation naturelle de la solution s'obtient en remplaçant les éléments différentiels par leurs accroissements. La discrétisation d'Euler de pas h de la diffusion $Y(t)$ solution de l'EDS précédente est le processus $(Y_{kh})_{k \in N}$ satisfaisant:

$$\begin{cases} Y_{(k+1)h} = Y_{kh} + hb(Y_{kh}) + \sqrt{h}\sigma(Y_{kh})z_{k+1}^{(h)} \\ Y(0) = y_0 \end{cases}$$

où les v.a $(z_k)_{k \in N}^{(h)}$ sont des vecteurs gaussiens de R^d indépendants, centrés et de matrice de covariance la matrice identité de R^d .

Ce processus Y_{kh} est une chaîne de Markov qui vérifie:

$$\begin{cases} h^{-1}E[(Y_{(k+1)h} - Y_{kh}) \mid Y_{kh} = y] = b(y) \\ h^{-1}Var[(Y_{(k+1)h} - Y_{kh}) \mid Y_{kh} = y] = \sigma(y)\sigma(y)' \end{cases}$$

Une telle chaîne de Markov peut se prolonger en un processus en temps continu en supposant qu'elle reste constante entre deux observations successives:

$$Y_t^{(h)} = Y_{kh}^{(h)} \text{ si } kh \leq t < (k+1)h \quad (k \in \mathbb{N})$$

Le processus $Y_t^{(h)}$ est continu à droite et limité à gauche. Pour l'étude de la convergence de ces processus, l'approche la plus simple consiste à étudier la convergence des lois fini-dimensionnelles des processus, mais il est aussi possible d'étudier la convergence des lois des processus $Y_t^{(h)}$ considérés globalement sur $[0, T]$ en se plaçant sur l'espace des fonctions càd-làg (fonctions continues à droite et limitées à gauche) muni d'une topologie adéquate: *la topologie de Skorohod* (Billingsley (1968)). Les résultats de convergence que nous donnons ici émanent des livres de Stroock-Varadhan (1979) et de Ethier-Kurtz (1986).

Soit une chaîne de Markov. Tel qu'on a vu dans le chapitre précédent, le modèle le plus simple pour modéliser le cours $S(t)$ d'un actif est le mouvement brownien géométrique:

$$dS(t) = S(t)(bdt + \sigma dW_t)$$

où b et σ sont des constantes réelles.

La solution $S(t)$ est strictement positive si la valeur initiale l'est et, en posant $Y(t) = \ln S(t)$, nous avons d'après la formule d'Itô:

$$dY(t) = (b - \frac{\sigma^2}{2})dt + \sigma dW_t.$$

La discréétisation d'Euler de pas h de ce modèle s'écrit:

$$Y_{kh} - Y_{(k-1)h} = (b - \frac{\sigma^2}{2})h + \sigma\sqrt{h}z_k$$

où les $(z_k)_{k \in \mathbb{N}}$ sont des gaussiennes indépendantes centrées réduites.

Ce modèle discret suppose la volatilité σ constante. Nous savons que cette hypothèse est peu conforme à la réalité et nous allons introduire dans ce dernier modèle, une variance conditionnelle non constante de type GARCH.

Le modèle que nous allons considérer sera donc le suivant:

$$Y_t = Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t$$

où les v.a $(z_t)_{t \in \mathbb{N}}$ sont i.i.d⁷, centrées de variance 1 (non nécessairement gaussiennes) et f est une fonction continue à valeurs réelles. Pour tout t de \mathbb{N} , la v.a σ_t dépendra du passé du processus z jusqu'à l'instant $t - 1$.

4.5 Conclusion

Il est maintenant connu que la volatilité d'un actif change de manière aléatoire au cours du temps. Pour décrire la volatilité dans le cas stochastique, la méthode la plus utilisée est le modèle ARCH (Auto Regressive Conditionally

⁷indépendantes et identiquement distribuées

Heteroscedastic). Le modèle ARCH spécifie la volatilité comme une fonction non-stochastique des observations passées. A cause de cette spécification, il est plus évident de calculer la vraisemblance du modèle ARCH et par conséquent estimer les paramètres en utilisant la procédure du maximum de vraisemblance. Une approche différente consiste à considérer un modèle à volatilité stochastique(SV), où le logarithme de la volatilité est considéré comme un processus stochastique linéaire tel qu'un processus ARMA. À la différence du modèle ARCH, il est difficile d'évaluer la vraisemblance exacte d'un modèle à volatilité stochastique et par conséquent leurs applications sont limitées. Les modèles SV constituent une discrétisation des modèles continus tels que le processus d'Ornstein-Uhlenbeck.

La méthode du maximum de vraisemblance (ML) n'est pas facile à implémenter dans le cas des modèles SV, mais il existe des méthodes alternatives. La méthode du Quasi-Maximum de vraisemblance(QML) est une de ces méthodes d'estimation développée par Nelson (1988), Harvey et *al.* (1994), et Ruiz (1994). Ils appliquent un filtre de Kalman standard à représentation linéaire de l'espace d'état du modèle SV pour obtenir une quasi-vraisemblance. Watanabe (1999) présente une méthode basée sur un filtre non-linéaire qui ne nécessite aucune transformation. Le filtre est décrit par une série d'intégrales qui ne peut pas être résolu analytiquement, mais Kitagawa (1997) a suggéré une méthode de décomposition en splines linéaires pour approximer le filtre. L'idée de base consiste à approximer les fonctions densités de probabilité appropriées pour chaque période par une fonction linéaire par morceaux.

Les modèles ARCH-GARCH ont une performance de prévision réduite, à

cause de la persistance de la variance excessive Avesani et Bertrand (1994). En particulier, on observe qu'un grand choc a des conséquences sur l'évolution future de la variance qui sont différentes de celles des petits chocs et que les grands chocs sont suivis par d'autres grands chocs, ce qui justifie l'effet du regroupement de la volatilité noté dans Mandelbrot (1963). Aussi, on remarque que les grands chocs impliquent une longue période de transition avant que le système ne retrouve son état naturel. C'est une des raisons pour lier l'analyse des processus stochastiques en temps continu avec la modélisation des séries financières. Nous n'avons pas réalisé d'études empiriques sur la modélisation ARCH-GARCH car ce n'est pas le but du présent mémoire. Nous renvoyons le lecteur intéressé à la modélisation ARCH en temps continu à l'article de Duan (1997) et à Engle (1995).

Chapitre 5

Modèles à Volatilité Stochastique

5.1 Introduction

La modélisation de la volatilité et son estimation jouent un rôle important dans l'évaluation des produits dérivés et dans les systèmes de gestion de risque. Il est maintenant reconnu que la simplicité du modèle de Black-Scholes, qui relie le prix d'une option au prix de l'actif sous-jacent et qui quantifie le risque à travers un paramètre de volatilité constante, n'est plus suffisant pour rendre compte des phénomènes des marchés modernes, spécialement depuis le crash de 1987. En effet, le modèle de Black-Scholes introduit un biais systématique dans les prix évalués, ce qu'on appelle l'effet *smile*. Nous savons maintenant que la série des rendements démontre un effet

leptokurtique, un coefficient d’asymétrie et une hétéroscédasticité conditionnelle dans la forme du regroupement de la volatilité(*volatility clustering*), lesquelles caractéristiques sont en désaccord avec les hypothèses de Black-Scholes. L’extension naturelle du modèle de Black-Scholes qui a été poursuivie dans la littérature et dans la pratique est de modifier la spécification de la volatilité, et ce en la considérant comme un processus stochastique. Ce qui rend cette approche particulièrement intéressante est premièrement que la volatilité est un processus caché, en ce sens qu’il permet l’évaluation des prix des options mais en même temps ce processus n’est pas observable. Deuxièmement, la volatilité a tendance à fluctuer à un haut niveau pendant un moment puis à fluctuer à un bas niveau pour une période similaire, et ainsi de suite. Sa moyenne se retourne plusieurs fois durant la vie d’un contrat contingent, ce qu’on appelle “*mean revert*”. Les contributions à la littérature sur les modèles à volatilité stochastique (SV) proviennent des deux domaines de recherche, les mathématiques financières et l’économétrie.

Le problème de la volatilité stochastique a été examiné par plusieurs auteurs dont Merton (1976), Geske (1979), Johnson (1987), Johnson et Shanno (1987), Eisenberg (1985), Wiggins (1987) et Scott (1987). Les articles de Merton et Geske présentent la solution à des types particuliers de problèmes à volatilité stochastique. Geske (1979) examine le cas où la volatilité de la valeur d’une firme est constante, telle que la volatilité du prix de l’option change de manière systématique quand le prix de l’option augmente ou diminue. Merton (1976) a examiné le cas où le prix suit un processus de diffusion avec saut. Johnson (1987) a étudié le cas général où la variance instantanée du prix de l’option suit un quelconque processus stochastique.

Cependant, dans le but de déduire les équations différentielles que le prix de l'option doit vérifier, il a supposé l'existence d'un actif financier avec un prix tel qu'il soit parfaitement corrélé instantanément avec la variance stochastique. L'existence d'un tel actif est suffisante pour déduire l'équation différentielle, mais Johnson a été incapable de la résoudre pour déterminer le prix de l'option.

Johson et Shanno (1987) ont obtenu quelques résultats numériques en utilisant des simulations, et produire un argument visant à expliquer le biais observé par Rubenstein (1985). Eisenberg (1985) a examiné comment les options doivent être évaluer relativement à chacune des autres options existantes en utilisant de pures arguments d'arbitrage.

Hull et White (1987) ont utilisé l'équation différentielle de Garman (1983) pour déduire une solution sous forme serielle du prix d'une option d'achat européenne sur un actif avec une volatilité qui n'est pas corrélée avec le prix de l'actif sous-jacent. Ils ont montré que pour un tel actif, le prix déduit de la formule de Black-Scholes sur-évalue les options en jeu ("*at-the-money*") et sous-évalue les options "*in*" et "*out-of-the-money*". Ils ont montré que la formule de Black-Scholes donne une bonne évaluation dans le cas où le prix de l'option se trouve dans un intervalle de 10% du prix d'exercice. Ils ont par ailleurs examiné le cas où la volatilité est corrélée avec le prix de l'option en utilisant des méthodes numériques. Les résultats ont montré que lorsque la corrélation est positive entre le prix de l'actif et sa volatilité, les options hors jeu (*out-of-the-money*) sont sous-évaluées par la formule de Black-Scholes, tandis que les options en jeu (*in-the-money*) restent sur-évaluées. Quand la

corrélation est négative, l'effet est inverse.

La discréétisation d'un modèle à volatilité stochastique introduit ce qu'on appelle variance autorégressive aléatoire (*ARV: Autoregressive Random Variance*). L'évolution d'un processus bi-dimensionnel en temps discret peut être décrite par la relation récurrente:

$$\begin{cases} \log S_t = \log S_{t-1} + \mu + \sigma_{t-1} \xi_t; \\ \log \sigma_t = \nu - \kappa(\nu - \ln \sigma_{t-1}) + \theta \eta_t \end{cases}$$

où (ξ_t, η_t) , $t \in N$, sont des variables aléatoires indépendantes, identiquement distribuées avec une loi gaussienne. Les modèles ARV ont été examiné, entre autres, par Viggins (1987), Chesney et Scott (1989), Melino et Turnbull (1990) ainsi que Duffie et Singleton (1993).

Une autre approche pour la modélisation de la volatilité stochastique dans un contexte discret, qui va être discutée plus en détail dans le prochain chapitre, est le modèle Autoregressive Conditionnal Heteroskedasticity(ARCH).

Un grand nombre de modèles en temps continu à volatilité stochastique a été proposé dans la littérature sous l'hypothèse naturelle que l'échec du modèle de Black-Scholes est attribuable au modèle avec mouvement brownien géométrique considéré pour représenter la dynamique du prix d'un actif. Un marché complet n'est pas, en général, atteignable dans le cas d'un modèle à volatilité stochastique. Hobson et Rogers (1998) ont suggéré une méthodologie qui conduit à un marché complet.

Jusqu'à récemment, il y avait une certaine dichotomie entre les modèles en temps discret, utilisés dans les travaux empiriques pour des séries chronologique financières et économiques, et les modèles en temps continu utilisés sou-

vent dans des travaux théoriques pour l'évaluation d'un actif. Il existe une littérature grandissante dévouée à réduire cette différence en considérant les modèles en temps continu obtenus quand le temps d'échantillonnage tend vers zéro avec diverses applications aux données haute fréquence, tel qu'illustré dans les travaux de Rossi (1996). L'estimation des paramètres d'un processus de diffusion observé discrètement avec des états inobservés est un problème difficile pour lequel un nombre de solutions ont été proposés dans la littérature. Le problème fondamental est que les fonctions densité de transition, et par conséquent la fonction de vraisemblance, ne peuvent être généralement exprimées avec une forme précise. Pour les modèles univariés, Aït-Sahalia (1996) a proposé une méthode, où l'équation de rétro-propagation de Kolmogorov est utilisée pour extraire un estimateur semi-paramétrique de la fonction de diffusion quand la fonction des cours (*drift*) est connue. En utilisant la même structure, Stanton (1997) a proposé une méthode semi-paramétrique pour l'estimation des approximations en temps discret du drift et de la fonction de diffusion. Les travaux de Aït-Sahalia (1996) ont été étudiés de manière approfondie par Chapman et Pearson (2000), où on montre que ces méthodes à base de noyaux produisent une fausse non linéarité. Malheureusement, il est difficile d'étendre ces méthodes aux cas des processus de diffusion multivariés, en particulier pour des processus avec certains états inobservés. Une autre branche de la littérature a étendu l'approche de Pederson (1995) qui considère le problème d'estimation comme un problème avec des valeurs manquantes en utilisant la méthode des chaînes de Markov Monté Carlo. Une troisième branche de la littérature s'intéresse aux méthodes basées sur les simulations et la comparaison des mo-

ments. La méthode des moments simulés de Duffie et Singleton (1993) fournit des conditions sur les moments en comparant les moments échantillonné avec les moments simulés à partir du modèle proposé. L'idée de base de la méthode d'inférence indirecte présentée dans les travaux de Smith (1993) et ceux de Gourieroux, Monfort et Renault (1993) est de comparer les moments simulés d'un modèle auxiliaire à temps discret en utilisant les données réelles avec les moments d'une version discrétisée du modèle en temps continu. Cette approche est utilisée pour un modèle à volatilité stochastique par Engle et Lee (1996). La méthode des moments efficients de Gallant et Tauchen (1996) est une méthode très similaire, mais plus affinée. Cette méthode a été utilisé pour un modèle à volatilité stochastique par Gallant, Hsieh et Tauchen (1997) et dans Gallant, Hsu et Tauchen (1998). L'inconvénient de la méthode générale des moments efficients(EMM) est qu'il requiert un modèle auxiliaire et un autre en temps continu qui capture toutes les caractéristiques des données et une comparaison un-à-un entre les paramètres de ces modèles.

On observe aussi un intérêt grandissant pour les méthode de filtrage dans la littérature de l'ingénierie financière: Les méthodes de filtrage pour les modèles structuraux en temps discret sont étudiés dans Harvey (1989) et sont appliqués aux modèles à volatilité stochastique en temps discret dans Harvey, Ruiz et Shepard (1993). Le filtrage de la volatilité à partir des prix des actifs dans un contexte ARCH et présenté dans les travaux de Nelson (1992, 1996). Timmer et Weigend (1997) traitent de l'estimation de la volatilité dans un modèle structurel linéaire, et ce en utilisant un filtre de Kalman ordinaire. Pour le systèmes non linéaires, le filtre de Kalman étendu(Extended Kalman Filter) fournit une solution approximative au problème de filtrage

(e.g Jaswinski (1970)), mais pour des équations différentielles stochastiques avec des fonctions de diffusion à états dépendants, nous avons besoin de filtre d'ordre supérieur, (Maybeck 1982). Un méthode de maximum de vraisemblance pour une estimation directe des paramètres embarqués dans les EDS est proposée par Nielsen et Madsen (2000), qui couvre aussi une généralisation des transformations proposées par Baadsgaard, Nielsen, Spliid, Madsen et Preisel (1997) tel que le filtre de Kalman étendu peut être appliqué à une classe d'EDS avec un terme de diffusion à états dépendant. Cependant, les modèles à volatilité stochastique ne correspondent pas à cette classe car nous devons utiliser des filtres d'ordre supérieur. Nielsen, Vestergaard et Madsen (2000) ont démontré que les méthodes de filtrage non linéaires permettent l'estimation d'états inobservés dans une grande classe de modèles à temps continu.

Dans ce chapitre, nous allons présenter dans un premier temps quelques caractéristiques de la volatilité que les modèles à volatilité stochastiques cherchent à reproduire et quelques motivations pour le choix de tels modèles. Nous ferons par la suite une présentation des modèles ARCH et de quelques modèles dérivés du modèle ARCH, tels que le modèle GARCH et leur utilisation dans la modélisation des actifs financiers ainsi que leurs cas limites. Dans la quatrième section, nous présenterons une méthodes d'estimation des modèles à volatilité stochastique par application d'une méthode de filtrage non-linéaire. Dans la cinquième section, nous présenterons une amélioration de cette même méthode de filtrage ainsi que les résultats obtenus. Nous terminerons ce chapitre par une discussion des résultats et une conclusion.

5.2 Quelques faits caractéristiques

La recherche d'un modèle de spécification est souvent guidée par des faits caractéristiques empiriques. La capacité d'un modèle de reproduire de tels faits est une caractéristique souhaitable et son incapacité de la faire est souvent un critère d'élimination d'une spécification , même si on n'essaye pas de reproduire ou d'expliquer toutes les régularités empiriques possibles en même temps avec un seul modèle. Dans cette section, nous présentons un résumé de faits caractéristiques concernant la volatilité.

5.2.1 Queues épaisses “(thick tails)”

Ce fait a été observé depuis le début des années soixante, notamment par Mandelbrot (1963), Fama (1963,1965) parmi d'autres auteurs qui ont montré que les fonctions des rendements(*asset return*) suivaient une distribution leptokurtique. Depuis, plusieurs articles ont proposé de modéliser la fonction des rendements comme variables indépendantes et identiquement distribuées, telle que la distribution ait des queues minces comme la loi de Pareto ou de Lévy.

5.2.2 Regroupement de la volatilité “ (*Volatility clustering*) ”

Toute observation fortuite d'une série chronologique financière révèle un groupe d'épisodes de haute et basse volatilité. En effet, le regroupement de la volatilité et les queues de distributions épaisses sont intimement liés. En effet, cette dernière caractéristique est une explication d'ordre statique tandis que les modèles ARCH fournissent un lien formel entre le comportement dynamique(conditionnel) et la présence de queues épaisses(inconditionnellement). Le modèle ARCH, introduit par Engle (1982) et ses nombreuses extensions survenues par la suite, comme les modèles à volatilité stochastique sont essentiellement construits pour imiter ce regroupement de la volatilité.

5.2.3 Effets de puissance “ (*Leverage effects*) ”

Un phénomène inventé par Black (1976) sous le nom d'effet de puissance suggère que le mouvement du prix d'un actif est négativement corrélé avec la volatilité. Donc la baisse du prix de l'actif implique une puissance rehaussée de la firme, ce qui demande cependant une plus grande incertitude et donc une plus grande volatilité. Des études empiriques rapportées par Black (1976), Chirtie (1982), et Schwert (1989) suggèrent, cependant, que cette puissance à elle seule n'est pas suffisante pour expliquer les asymétries qu'on observe dans la courbe des prix d'un actif. D'autres études concernant l'effet de puissance ont été menées par Nelson (1991), Gallant, Rossi et Tauchen (1992,1993), Campbell et Kyle (1993), ainsi que Engle et Ng (1993).

5.2.4 Arrivée d'information

Les rendements sont typiquement mesurés et modélisés avec des observations échantillonnées avec des fréquences fixes, tels que quotidiennement, hebdomadairement ou des observations mensuelles. Plusieurs auteurs, dont Mandelbrot et Taylor (1967) et Clark (1973), ont suggéré de lier cette fonction explicitement avec le flot des informations arrivées. L'arrivée d'information n'est pas uniforme dans le temps et souvent n'est pas directement observable. Conceptuellement, on peut considérer le mouvement du prix d'un actif comme une réalisation d'un processus $Y_t = Y_{Z_t}^*$, où Z_t est dit processus de direction. Ce processus non décroissant Z_t peut être vu comme relié à l'arrivée d'information. Cette idée de la déformation temporelle ou de processus stochastique subordonné a été utilisée par Mandelbrot et Taylor (1967) afin d'expliquer la forme des queues épaisses, par Clark (1973) pour expliquer la volatilité et a été récemment raffiné et exploré par Ghysels, Gouriéroux et Jasiak (1995a). D'ailleurs, Easley et O'Hara (1992) ont présenté un modèle de microstructure utilisant cette déformation temporelle. En pratique, elle suggère un lien direct entre la volatilité du marché et (1) le volume transigé, (2) l'arrivée des cotes, (3) les événements prévisibles tels que l'annonce de distribution de dividendes ou les données macroéconomiques disponibles, (4) la fermeture des marchés, parmi d'autres phénomènes liés à l'arrivée d'informations nouvelles sur le marché.

Structure de la volatilité implicite et l'effet “Smile”: Le modèle de Black-Scholes prédit un terme constant de la volatilité. En réalité la structure à terme de la volatilité implicite des options en jeu (at-the-money) possède une

inclinaison ascendante quand la volatilité est faible et inversement quand la volatilité est élevée (voir Stein (1989)). Taylor et Xu (1994) ont montré que la structure de la volatilité implicite des options renverse sa pente avec une fréquence constante de quelques mois.

5.3 Modèles bivariés à volatilité stochastique

Dans le présent paragraphe, on présente un aperçu des processus de diffusion existant qui modélisent la volatilité stochastique. Tous ces modèles cherchent à étendre celui de Black-Scholes en considérant la volatilité comme processus de diffusion:

$$\begin{cases} dS_t = \alpha S_t dt + \sigma_t S_t dW_{1,t}; \\ d\psi(\sigma_t) = a(\sigma_t) dt + b(\sigma_t) dW_{2,t} \end{cases}$$

où $W_{1,t}$ et $W_{2,t}$ sont deux processus de Wiener corrélés avec un coefficient de corrélation ρ . Le taux de rendement instantané est α , $\psi(\sigma_t)$ est une fonction de σ_t (comme $\log(\sigma_t, \sigma_t^2, \sqrt{\sigma_t}, \text{etc})$), $a(\sigma_t)$ et $b(\sigma_t)$ représentent respectivement la moyenne et la variance instantanée de la volatilité stochastique σ_t . Le coefficient de corrélation ρ représente l'effet de levage, le fait que la hausse de la volatilité ait un moindre impact sur l'évolution du processus que la baisse d'une même amplitude de la volatilité. Le tableau 5.1 présente un nombre de spécifications données dans la littérature.

En pratique, les observations réelles Y_k , sont les valeurs mesurées aux instants $t_1 < \dots < t_i < t_N$ où N est le nombre d'observations. L'équation

$\psi(\sigma_t)$	$a(\sigma_t)$	$b(\sigma_t)$	Références
σ_t	$\kappa(\beta - \sigma_t)$	θ	[Stein et Stein, 1991]
σ_t	$\kappa(\beta - \sigma_t)$	$\theta\sigma_t$	[Courtadon, 1982]
σ_t	$\kappa\sigma_t(\beta - \sigma_t)$	$\theta\sigma_t$	[Scott, 1987]
σ_t	$\kappa\sigma_t$	$\theta\sigma_t$	[Hull et White, 1987a][Johson et Shanno, 1987]
σ_t	$\frac{(\nu - \kappa\sigma_t^2)}{\sigma_t}$	θ	[Hull et White, 1988] [Baily et Stulz, 1989][Heston, 1993]
σ_t	$\kappa(\beta - \sigma_t^2)$	$\theta\sigma_t$	[Musiela et Rutkowski, 1997]
$ln\sigma_t$	$\kappa(\beta - ln\sigma_t)$	θ	[Wiggins, 1987][Scott, 1991]

Tableau 5.1: Aperçu des différentes modélisation de la volatilité stochastique.

des observations est la suivante:

$$Y_k = S_{t_k} + e_k, k = 1, \dots, N$$

où les e_k forment un processus de bruit blanc gaussien de moyenne 0 et de variance σ_e^2 . Les processus stochastiques $W_t = (W_{1,t}, W_{2,t})$ et e_k sont supposés être indépendants pour tout t et k . Comme on le voit dans le tableau précédent, un grand nombre de modèles à volatilité stochastique ont été proposés dans la littérature. Les équations (4.1) et (4.2) permettent de spécifier et d'estimer un modèle à volatilité stochastique de manière générale. En particulier, une estimation de la volatilité stochastique σ_t et des paramètres de l'EDS (4.1) peuvent être obtenus.

Chapitre 6

Méthodes de filtrage non linéaire

Ce chapitre traite essentiellement du filtrage linéaire et non linéaire dans l'estimation de l'état d'un système et les aspects théoriques sur la prédiction dans les modèles d'espaces d'états. Il existe deux parties importantes dans la théorie du filtrage, les filtres exacts et les filtres approximatifs, respectivement. Dans les filtres exacts, on utilise une information généralement plus complète sur la distribution des observations, tandis que dans les filtres approximatifs, on introduit des hypothèses simplificatrices sur la distribution des observations. Dans les modèles linéaires, le filtre de Kalman donne une solution exacte du problème de filtrage. Dans le cas non linéaire, le filtre de Kalman étendu [EKF], Harvey (1989) fournit une solution approximative: le filtre de Kalman étendu utilise l'estimation de l'état présent dans une

linéarisation pour ensuite utiliser le filtre de Kalman ordinaire au problème linéarisé obtenu. Dans le modèle à états dépendants, le filtrage implique l'équation de rétro-propagation de Kolmogorov et la formule de Bayes. Afin d'éviter les calculs complexes qu'impliquent l'équation de rétro-propagation de Kolmogorov et de la formule de Bayes, et afin d'obtenir un algorithme computationnellement réalisable, certaines hypothèses sont faites sur la densité conditionnelle.

Selon différentes suppositions simplificatrices, nous présenterons le filtre du second ordre tronqué (TSOF) ou le filtre gaussien du second ordre (GSOF).

6.1 Description du problème de filtrage de manière générale

La recherche en filtrage est motivée par de nombreuses applications dans des domaines variés. Ce problème est en effet assez général puisqu'il s'agit d'estimer l'état du système dynamique à partir d'observations bruitées. Nous décrivons ici brièvement le contexte d'application du filtrage.

Les systèmes physiques ou économiques sont souvent sujets à des perturbations aléatoires. La modélisation de l'évolution dans le temps des états $(X_t)_{t \geq 0}$ du système considéré s'écrit alors sous la forme d'un terme d'évolution déterministe et d'un terme aléatoire,

$$X_t = X_0 + \int_0^t f(s, X_s) ds + \int_0^t b(X_s) dW_s, \quad X_0 \text{ de loi donnée,} \quad (6.1)$$

où f est une fonction connue caractérisant la partie déterministe de la dynamique et $(W_t)_{t \geq 0}$ le processus de Wiener standard modélisant les perturbations aléatoires de la dynamique et l'incertitude sur le choix de la fonction f . Cette équation de dynamique traduit la connaissance à priori que l'on a du système. Dans le but de déterminer l'état précis du système, on élabore une procédure d'observation permettant de recueillir, à des instants discrets t_n , des mesures des états du système.

Ces observations sont, en général, entachées d'erreurs dues à l'imperfection de l'instrument de mesure. La suite des observations $(Y_n)_{n \geq 1}$ est modélisée par l'équation suivante, reliant l'observation Y_n à l'état courant du système X_{t_n} ,

$$Y_n = h_n(X_{t_n}) + V_n, \quad (6.2)$$

où h_n est la fonction d'observation connue et $(V_n)_{n \geq 1}$ une suite de variables aléatoires de statistique connue modélisant l'imperfection des observations.

Le problème consistant à déterminer l'état du système, X_{t_n} , à partir des observations bruitées (Y_1, \dots, Y_k) est appelé estimation:

- si $k > n$, on parle plus précisément de lissage,
- si $k = n$, on parle de filtrage,
- si $k < n$, on parle de prédiction.

Le choix du critère qualifiant la précision de l'estimée n'est pas unique. Si l'on choisit la distance moyenne quadratique, il est bien connu que l'estimée

optimale de X_{t_n} , au vu des observations (Y_1, \dots, Y_k) est l'espérance conditionnelle:

$$\mathbb{E}[X_{t_n} | Y_1, \dots, Y_k].$$

Cependant, cet estimateur n'est pas toujours approprié. Typiquement, lorsque la loi conditionnelle de X_{t_n} sachant les observations (Y_1, \dots, Y_k) est fortement multimodale, on lui préfère le maximum a posteriori (*maximisant la loi conditionnelle $\mathcal{L}(X_{t_n} | Y_1, \dots, Y_k)$*). Dans tous les cas, l'information la plus riche disponible au vu des observations (Y_1, \dots, Y_k) est contenue dans la loi conditionnelle de X_{t_n} sachant les observations passées. Avec la donnée de cette loi, on est libre dans le choix de l'estimateur. Le problème de filtrage consistera donc à calculer (de manière exacte ou approchée) cette mesure de probabilité conditionnelle. En toute généralité, nous avons affaire à un problème de dimension infinie.

6.2 Le modèle

Dans le cas le plus général, on considère que le système est régi par le système d'équations différentielles stochastiques suivantes:

$$d\mathbf{X}_t = \mathbf{f}(\mathbf{X}_t, t)dt + \mathbf{g}(\mathbf{X}_t, t)d\mathbf{W}_t, \mathbf{X}_0 = \mathbf{x}, t > t_0 \quad (6.3)$$

où $\mathbf{W}_t, t \geq t_0$ est un vecteur de processus de Wiener de variance incrémentale \mathbf{Q}_t ¹ et de dimension $s \in \mathbb{N}_+$. On notera que dans le cas d'un espace à

¹On inclut \mathbf{Q}_t en général dans la fonction g

états non dépendants, \mathbf{g} et \mathbf{X}_t sont indépendants. Les observations sont échantillonnées à chaque instant discret t_k tel que:

$$Y_k = h(X_{t_k}, t_k, \theta) + e_k, \quad (6.4)$$

où $e_k \in N_m(0, V_k)$, h est un vecteur de fonctions de X_{t_k}, t_k et θ , où θ est un vecteur paramètre.

6.3 Filtrage discret-discret

6.3.1 Filtre de Kalman

La méthode de Kalman est une procédure d'estimation dynamique des paramètres qui sont fonctions du temps. Le filtre de Kalman est utilisé pour estimer l'état $x \in \mathbb{R}^n$ d'un système dynamique défini par une équation linéaire:

$$x_k = A_k x_{k-1} + B_k u_k + w_{k-1},$$

avec une mesure $z \in \mathbb{R}^m$:

$$z_k = H_k x_k + v_k,$$

$A_{n \times n}$ est une matrice qui fait le lien entre les paramètres du système à deux étapes successives. $B_{n \times l}$ fait le lien entre les valeurs optionnelles de contrôles et l'état du système. $H_{m \times n}$ est la matrice qui fait le lien entre les paramètres du système et les mesures. w_k et v_k sont respectivement des bruits sur des paramètres à estimer et sur des mesures (observations). w_k et v_k sont des variables aléatoires dont la distribution est normale de moyennes nulles et de

matrices de covariances respectives Q_k et R_k . Donc:

$$w_k \sim N(0, Q_k)$$

$$v_k \sim N(0, R_k)$$

Pour utiliser le filtre de Kalman, les matrices de covariance Q_k et R_k doivent être évaluées à l'avance. La détermination de Q_k est généralement plus difficile parce que l'observation directe de l'état du système est impossible. Il y a deux étapes principales pour l'évaluation des paramètres du système; l'évaluation *a priori* (*la prédition*) et l'évaluation *a posteriori* (*la correction*). Les paramètres du système sont estimés premièrement à partir de valeurs de l'étape précédente et en suite ils sont corrigés par des mesures dans l'étape de correction.

On définit $\hat{x}_k^- \in \mathbb{R}$ comme étant l'estimation *a priori* à l'état k sachant les mesures jusqu'à l'instant $k-1$, et $\hat{x}_k \in \mathbb{R}$ l'estimation *a posteriori* du système à l'état k sachant la mesure z_k . On peut donc définir les erreurs d'estimation *a priori* et *a posteriori* par:

$$e_k^- = x_k - \hat{x}_k^- \text{ et,}$$

$$e_k = x_k - \hat{x}_k.$$

L'estimation *a priori* et *a posteriori* de la covariance de l'erreur est donnée respectivement par:

$$P_k^- = E[e_k^- e_k^{-\top}] \text{ et,}$$

$$P_k = E[e_k e_k^\top]. \quad (6.5)$$

Pour définir les équations du filtre de Kalman, nous commençons par définir une équation mettant en relation linéaire l'état du système *a posteriori* \hat{x}_k d'une part et l'estimation du système *a priori* \hat{x}_k^- et la différence entre la mesure actuelle z_k et la mesure $H_k \hat{x}_k^-$ prédite à partir de l'état z_{k-1} d'autre part, tel que:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-)$$

La différence $(z_k - H_k \hat{x}_k^-)$ est dite *l'innovation* de la mesure. La matrice K_k est choisie pour être le facteur de gain qui minimise la covariance de l'erreur *a posteriori*. Une des formes de K qui minimise l'équation (6.5) est

$$K_k = P_k^- H_k^\top (H_k P_k^- H_k^\top + R_k)^{-1}$$

Les équations du filtre de Kalman peuvent être divisées en deux groupes: les équations de *mise à jour temporelle*, et les équations de *mise à jour des mesures*. Les équations de mise à jour temporelle sont responsables de la projection avant (projection temporelle) de l'état courant du système et de l'estimation de la covariance de l'erreur pour obtenir une estimation *a priori* du système à l'état suivant. Les équations de mise à jour des mesures servent à incorporer une nouvelle mesure dans l'estimation *a priori* pour obtenir une estimation *a posteriori* améliorée. Les équations spécifiques pour la mise à jour temporelle sont,

$$\hat{x}_{k+1}^- = A_k \hat{x}_k^+ B_k u_k,$$

$$P_{k+1}^- = A_k P_k A_k^\top + Q_k.$$

Voici par ailleurs les équations spécifiques pour la mise à jour des mesures sont,

$$\begin{aligned} K_k &= P_k^- H_k^\top (H_k P_k^- H_k^\top + R_k)^{-1}, \\ \hat{x}_k &= \hat{x}_k^- + K(z_k - H_k \hat{x}_k^-), \\ P_k &= (I - K_k H_k) P_k^-. \end{aligned}$$

Notons finalement que si Q_k et R_k sont constants, l'estimation de la covariance de l'erreur P_k et le gain de Kalman K_k vont se stabiliser rapidement². Le filtre de Kalman peut être vu comme un algorithme de prédition-correction pour la résolution de problèmes numériques.

6.3.2 Méthode du Quasi-Maximum de vraisemblance

On suppose $\{y_t\}$ décrit par le processus stochastique suivant:

$$y_t = \sigma_t u_t; \quad u_t \sim NID(0, 1)^3 \quad (6.6)$$

où σ_t est la volatilité de y_t .

Les modèles à volatilité stochastique (SV) spécifient le logarithme de la volatilité comme étant un processus linéaire stochastique. La version la plus

²Pour plus de détails sur le filtre de Kalman, voir les articles de Maybeck (1979), Brown (1992) et celui de Jacobs (1993).

³De loi normale $N(0,1)$ et identiquement distribué.

simple, le modèle AR(1) est utilisée:

$$\log(\sigma_t^2) = \alpha + \beta \log(\sigma_{t-1}^2) + \omega_t, \quad \alpha > 0, 0 \leq \beta < 1, \quad \omega_t \sim NID(0, \sigma_\omega^2) \quad (6.7)$$

où u_t et ω_s sont indépendants deux à deux pour tout t et s .

Soit θ le vecteur des paramètres inconnus $(\alpha, \beta, \sigma_\omega)$. La vraisemblance est définie par :

$$\begin{aligned} L &= p(y_1, \dots, y_T | \theta) \\ &= \int \left[\prod_1^T \left\{ \frac{\exp\{-y_t^2/2\sigma_t^2\}}{\sqrt{2\pi\sigma_t^2}} \right\} \right] \times p(\sigma_1, \dots, \sigma_T | \theta) d\sigma_1 d\sigma_2 \dots d\sigma_T \end{aligned} \quad (6.8)$$

où $p(\cdot | \cdot)$ représente la densité conditionnelle et T le nombre d'observations. Cette intégrale ne peut pas être résolue de manière analytique. La méthode du Quasi-Maximum de vraisemblance (QML) est une des méthodes qui permet de contourner ce problème. Cette méthode a été développée par Nelson (1988), Harvey et *al* (1994), et Ruiz (1994). La méthode QML transforme d'abord le modèle à volatilité stochastique (SV) en forme d'espace d'état linéaire:

$$\log(y_t^2) = \log(\sigma_t^2) + \log(u_t^2) \quad (6.9)$$

On définit $Y_t = \log(y_t^2)$; $x_t = \log(\sigma_t^2)$, et $\nu_t = \log(u_t^2)$. Ce qui revient donc au système d'équations suivant:

$$\begin{cases} Y_t = x_t + \nu_t \\ x_t = \alpha + \beta x_{t-1} + \omega_t \end{cases}$$

où ν_t n'est pas de loi normale, sa moyenne est égale à $E(\nu_t) = 1.27$.

Par la suite un filtre de Kalman⁴ standard est appliqué au modèle. Les équations du filtre de Kalman sont:

Prédiction du premier pas:

$$\begin{cases} x_{t|t-1} = \alpha + \beta x_{t-1|t-1} \\ P_{t|t-1} = \beta^2 P_{t-1|t-1} + \sigma_\omega^2 \end{cases}$$

Mise-à-jour:

$$\begin{cases} x_{t|t} = x_{t|t-1} + \frac{P_{t|t-1}}{f_t} v_t \\ P_{t|t} = P_{t|t-1} - \frac{P_{t|t-1}^2}{f_t} \end{cases}$$

où v_t et f_t sont respectivement l'erreur de prédiction et sa variance, telles que:

$$\begin{cases} v_t = Y_t - x_{t|t-1} + 1.27 \\ f_t = P_{t|t-1} + \frac{\pi^2}{2} \end{cases}$$

Si l'on connaît les valeurs initiales $x_{1|0}$ et $P_{1|0}$, on peut résoudre ce filtre récursivement. D'après Watanabe (1999), si $|\beta| < 1$, la moyenne inconditionnelle et la variance de x_t sont données respectivement par: $\alpha/(1 - \beta)$ et $\sigma_\omega^2/(1 - \beta^2)$. Ces dernières valeurs sont généralement utilisées pour $x_{1|0}$ et $P_{1|0}$

Si les termes d'erreurs v_t et ω_t avaient une distribution normale, la log-vraisemblance serait:

$$\log L = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(f_t) - \frac{1}{2} \sum_{t=1}^T \frac{v_t^2}{f_t} \quad (6.10)$$

⁴Voir Abramovitk et Stegan (1970). Nous renvoyons le lecteur intéressé par le filtre de Kalman à l'article de Harvey (1989)

où T désigne le nombre d'observations.

On pourrait par la suite optimiser cette fonction de vraisemblance pour obtenir les paramètres $(\alpha, \beta, \sigma_\omega)$ de notre modèle SV. Malheureusement, telle n'est pas le cas car v_t n'est pas normale. Donc, cette méthode ne se base pas sur la vraie vraisemblance, et c'est pour cela qu'on l'appelle Estimation Quasi-maximum de Vraisemblance(QML).

Cette estimation QML est facile à implémenter, mais cependant, selon l'évidence Monte-Carlo de Jacquier et *al* (1994), n'est pas performante pour des échantillons de taille finie car elle ne se base pas sur la vraie vraisemblance. Un autre problème de cette méthode, est que le modèle qu'on choisit doit pouvoir être transformable en forme d'espace d'état linéaire, ce qui restreint le nombre de ces modèles.

6.3.3 Méthode de filtrage Non-Linéaire NFML: Non Linear Filtering Maximum Likelihood

Cette méthode permet, entre autre, l'analyse des modèles SV, tel que celui défini par les équations (6.6) et (6.7), par application d'un filtre non-linéaire au modèle, sans aucune transformation préalable. Un tel filtre est basé uniquement sur le théorème de Bayes. A la différence du filtre de Kalman, qui est décrit par les seuls deux premiers moments, grâce aux propriétés des lois normales, le filtre proposé par Watanabe (1999) que nous présentons ici, est représenté en utilisant les fonctions densités de probabilité conditionnelles.

On définit $x_t \equiv \log(\sigma_t)$ et $\mathbf{y}_t \equiv (y_1, y_2, \dots, y_t)$.

Des équations (6.6) et (6.7), on a:

$$\begin{cases} p(x_t|x_{t-1}, \mathbf{y}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_\omega^2}} \exp\left[-\frac{(x_t - \alpha - \beta x_{t-1})^2}{2\sigma_\omega^2}\right] \\ p(y_t|x_t, \mathbf{y}_{t-1}) = \frac{1}{\sqrt{2\pi \exp(x_t)}} \exp\left[-\frac{y_t^2}{2\exp(x_t)}\right] \end{cases}$$

Les équations de ce filtre sont données par les formules suivantes:

Prédiction du premier pas:

$$\begin{aligned} p(x_t|\mathbf{y}_{t-1}) &= \int_{-\infty}^{\infty} p(x_t, x_{t-1}|\mathbf{y}_{t-1}) dx_{t-1} \\ &= \int_{-\infty}^{\infty} p(x_t|x_{t-1}, \mathbf{y}_{t-1}) p(x_{t-1}|\mathbf{y}_{t-1}) dx_{t-1} \end{aligned} \tag{6.11}$$

Mise-à-jour:

$$\begin{aligned} p(x_t|\mathbf{y}_t) &= p(x_t|y_t, \mathbf{y}_{t-1}) \\ &= \frac{p(x_t, y_t|\mathbf{y}_{t-1})}{p(y_t|\mathbf{y}_{t-1})} \\ &= \frac{p(y_t|x_t, \mathbf{y}_{t-1}) p(x_t|\mathbf{y}_{t-1})}{p(y_t|\mathbf{y}_{t-1})} \end{aligned} \tag{6.12}$$

où:

$$\begin{aligned} p(y_t|\mathbf{y}_{t-1}) &= \int_{-\infty}^{\infty} p(y_t, x_t|\mathbf{y}_{t-1}) dx_t \\ &= \int_{-\infty}^{\infty} p(y_t|x_t, \mathbf{y}_{t-1}) p(x_t|\mathbf{y}_{t-1}) dx_t \end{aligned} \tag{6.13}$$

Le problème de ce filtre est qu'il est difficile à résoudre de manière analytique. Kitagawa (1987) a proposé une technique de spline linéaire pour ap-

proximer ce filtre. L'idée de base de cette méthode est d'approximer les fonctions densités de probabilité pour chaque période par des fonctions linéaires par morceaux. En utilisant la règle du trapèze, on approxime les équations (6.11), (6.12) et (6.13) par:

Prédiction du premier pas:

$$\begin{aligned}
p(x_t^{(i)} | \mathbf{y}_{t-1}) &= \int_{-\infty}^{\infty} p(x_t^{(i)} | x_{t-1}, \mathbf{y}_{t-1}) p(x_{t-1} | \mathbf{y}_{t-1}) dx_{t-1} \\
&\approx \frac{1}{2} \sum_{n=1}^N (x_{t-1}^{(n)} - x_{t-1}^{(n-1)}) [p(x_t^{(i)} | x_{t-1}^{(n-1)}, \mathbf{y}_{t-1}) \\
&\quad + p(x_t^{(i)} | x_{t-1}^{(n)}, \mathbf{y}_{t-1}) p(x_{t-1}^{(n)} | \mathbf{y}_{t-1})]; i = 0, 1, \dots, N
\end{aligned} \tag{6.14}$$

Mise-à-jour:

$$p(x_t^{(i)} | \mathbf{y}_t) = \frac{p(y_t | x_t^{(i)}, \mathbf{y}_{t-1}) p(x_t^{(i)} | \mathbf{y}_{t-1})}{p(y_t | \mathbf{y}_{t-1})}; i = 0, 1, \dots, N \tag{6.15}$$

où:

$$\begin{aligned}
p(y_t | \mathbf{y}_{t-1}) &\approx \frac{1}{2} \sum_{n=1}^N (x_t^{(n)} - x_t^{(n-1)}) [p(y_t | x_t^{(n-1)}, \mathbf{y}_{t-1}) p(x_t^{(n-1)} | \mathbf{y}_{t-1}) \\
&\quad + p(y_t | x_t^{(n)}, \mathbf{y}_{t-1}) p(x_t^{(n)} | \mathbf{y}_{t-1})]; i = 0, 1, \dots, N
\end{aligned} \tag{6.16}$$

Une fois qu'on connaît la variable d'état initiale $p(x_1) (\equiv p(x_1 | \mathbf{y}_0))$, les équations (6.14), (6.15) et (6.16) peuvent être résolues récursivement. La distribution de x_t est normale avec pour moyenne $\alpha/(1 - \beta)$ et de variance $\sigma_\omega^2/(1 - \beta^2)$. Ces valeurs sont utilisées comme valeurs initiales des variables d'état.

La Log-vraisemblance est définie par:

$$\log L = \log[p(\mathbf{y}_T)] = \sum_{t=1}^T \log[p(y_t|\mathbf{y}_{t-1})].$$

Contrairement à la méthode d'estimation QML, cette méthode donne la log-vraisemblance exacte tant que la perte causée par l'approximation spline linéaire est négligeable. Les paramètres du modèle SV sont alors estimés en maximisant la log-vraisemblance exacte. On appelle cette méthode NFML: *Non-Linear Filtering Maximum Likelihood*.

Pour l'implémentation de cette méthode, nous devons choisir les noeuds $x_t^{(i)}$ pour $i = 0, 1, \dots, N$ et $t = 1, \dots, T$. Généralement, on utilise comme nombre de noeuds $N=25$ ou $N=50$ ⁵. La distribution des noeuds doit être similaire à la vraie distribution de x_t sachant \mathbf{y}_{t-1} et \mathbf{y}_t . Et comme cette dernière valeur change au cours du temps, nous devons permettre aux noeuds de changer au cours du temps. Pour ce faire, nous utilisons la procédure de Tanizaki (1993) pour le choix des noeuds: On estime d'abord les paramètres SV par la méthode QML, ensuite selon ses estimations, on exécute le filtre de Kalman pour obtenir $x_{t|t-1}, x_{t|t}, P_{t|t-1}$ et $P_{t|t}$ pour $t = 1, \dots, T$. Par la suite on génère $x_t^{(i)}$ de la fonction densité $\frac{1}{2}N(x_{t|t-1}, cP_t|t-1) + \frac{1}{2}N(x_{t|t}, cP_t|t)$, où $c=4$ selon Tanizaki (1993).

Un autre problème avec cette approximation linéaire par morceaux est que les erreurs computationnelles s'accumulent et parfois les approximations de $p(x_t|\mathbf{y}_{t-1})$ ou de $p(x_t|\mathbf{y}_t)$ ne donnent pas une intégrale de 1. Pour éviter cette

⁵Voir l'article de Watanabe (1999)

situation, on implémente un ajustement (*ad-hoc*), en multipliant par une constante $p(x_t^{(i)} | \mathbf{y}_{t-1})$ ($i=0,1,\dots,N$) et ce pour tout $t = 1, \dots, T$ de telle manière que

$$\frac{1}{2} \sum_{i=1}^N (x_t^{(i)} - x_t^{(i-1)}) [p(x_t^{(i-1)} | \mathbf{y}_{t-1}) + p(x_t^{(i)} | \mathbf{y}_{t-1})] = 1.$$

6.3.4 Comparaison des estimations

On considère l'estimation par lissage de la volatilité, σ_t . Les équations de l'algorithme standard de lissage sont:

$$x_{t|T} = x_{t|t} + P_t^*(x_{t+1|T} - x_{t+1|t}) \quad (6.17)$$

$$P_{t|T} = P_{t|t} + P_t^{*2}(P_{t+1|T} - P_{t+1|t}) \quad (6.18)$$

où

$$P_t^* = \beta \frac{P_{t|t}}{P_{t+1|t}}. \quad (6.19)$$

Si on exécute le filtre de Kalman conditionnellement aux estimations QML des paramètres SV on trouve $x_{t|t-1}, P_{t|t-1}, x_{t|t}$ et $P_{t|t}$ pour $t = 1, \dots, T$. Sachant ces valeurs, on peut résoudre les équations (6.17), (6.18), et (6.19). On ne connaît pas la vraie distribution de x_t conditionnellement aux informations à l'instant T, mais en l'approximant avec une normale de moyenne $x_{t|T}$ et de variance $P_{t|T}$, on obtient

$$\sigma_{t|T}^2 = \exp[x_{t|T} + \frac{1}{2}P_{t|T}] \quad (6.20)$$

On appelle cette estimation de la volatilité: *Estimation Lisse Standard de la Volatilité*. Cette méthode présente deux problèmes. Premièrement, l'estimateur de lissage $x_{t|T}$ est le MMLSE: *Minimum Mean Linear Square Estimate* et non le MMSE: *Minimum Mean Square Estimate* de x_t car l'erreur des observations v_t n'est pas de distribution normale. Deuxièmement, la vraie distribution de x_t conditionnellement aux informations à l'instant T n'est pas normale ce qui rend problématique l'utilisation de l'équation (6.20) pour l'estimation de la volatilité σ_t^2 .

Un algorithme de lissage peut être appliqué directement au modèle SV donné par les équations (6.6) et (6.7) sans aucune transformation et il est optimal dans le sens qu'il donne le MMSE de la volatilité σ_t^2 . Cet algorithme est dérivé du théorème de Bayes et il est représenté en utilisant les fonctions densités de probabilité conditionnelles.

On considère d'abord la densité jointe de x_t et x_{t+1} , conditionnellement aux informations à l'instant T;

$$\begin{aligned}
p(x_t, x_{t+1} | \mathbf{y}_T) &= p(x_{t+1} | \mathbf{y}_T) p(x_t | x_{t+1}, \mathbf{y}_T) \\
&= p(x_{t+1} | \mathbf{y}_T) p(x_t | x_{t+1}, \mathbf{y}_t) \\
&= \frac{p(x_{t+1} | \mathbf{y}_T) p(x_t, x_{t+1} | \mathbf{y}_t)}{p(x_{t+1} | \mathbf{y}_t)} \\
&= \frac{p(x_{t+1} | \mathbf{y}_T) p(x_{t+1} | x_t, \mathbf{y}_t) p(x_t | \mathbf{y}_t)}{p(x_{t+1} | \mathbf{y}_t)}
\end{aligned} \tag{6.21}$$

En intégrant cette quantité par rapport à x_{t+1} on a

$$\begin{aligned}
p(x_t | \mathbf{y}_T) &= \int_{-\infty}^{\infty} p(x_t, x_{t+1} | \mathbf{y}_T) dx_{t+1} \\
&= p(x_t | \mathbf{y}_t) \int_{-\infty}^{\infty} \frac{p(x_{t+1} | \mathbf{y}_T) p(x_{t+1} | x_t, \mathbf{y}_t)}{p(x_{t+1} | \mathbf{y}_t)} dx_{t+1}
\end{aligned} \tag{6.22}$$

C'est l'algorithme optimal de lissage pour le modèle SV. On obtient donc

$$E(\sigma_t^2 | \mathbf{y}_T) = \int_{-\infty}^{\infty} \exp(x_t) p(x_t | \mathbf{y}_T) dx_t. \tag{6.23}$$

Cette dernière quantité peut être approximée en utilisant une méthode de splines linéaires tel qu'on a vu dans le paragraphe précédent. En choisissant les noeuds convenablement, les équations (6.22) et (6.23) peuvent être alors approximées par les équations:

$$\begin{aligned}
p(x_t^{(i)} | \mathbf{y}_T) &\approx p(x_t^{(i)} | \mathbf{y}_t) \sum_{m=1}^M \int_{x_{t+1}^{(m+1)}}^{x_{t+1}^{(m)}} \frac{p(x_{t+1} | \mathbf{y}_T) p(x_{t+1} | x_t^{(i)}, \mathbf{y}_t)}{p(x_{t+1} | \mathbf{y}_t)} dx_{t+1} \\
&\approx \frac{p(x_t^{(i)} | \mathbf{y}_t)}{2} \sum_{m=1}^M (x_{t+1}^{(m)} - x_{t+1}^{(m-1)}) \\
&\times \left[\frac{p(x_{t+1}^{(m-1)} | \mathbf{y}_T) p(x_{t+1}^{(m-1)} | x_t^{(i)}, \mathbf{y}_t)}{p(x_{t+1}^{(m-1)} | \mathbf{y}_t)} + \frac{p(x_{t+1}^{(m)} | \mathbf{y}_T) p(x_{t+1}^{(m)} | x_t^{(i)}, \mathbf{y}_t)}{p(x_{t+1}^{(m)} | \mathbf{y}_t)} \right] \\
&; \text{ pour } i = 0, 1, \dots, M
\end{aligned} \tag{6.24}$$

et

$$E(\sigma_t^2 | \mathbf{y}_T) \approx \sum_{m=1}^M \int_{x_t^{(m-1)}}^{x_t^{(m)}} \exp(x_t) p(x_t | \mathbf{y}_T) dx_t$$

$$\begin{aligned}
&\approx \frac{1}{2} \sum_{m=1}^M (x_t^{(m)} - x_t^{(m-1)}) \left[\exp(x_t^{(m-1)}) p(x_t^{(m-1)} | \mathbf{y}_T) \right. \\
&\quad \left. + \exp(x_t^{(m)}) p(x_t^{(m)} | \mathbf{y}_T) \right]
\end{aligned} \tag{6.25}$$

Pour le choix des noeuds, on estime d'abord les paramètres SV par la méthode QML, ensuite selon ses estimations, on exécute le filtre de Kalman pour obtenir $x_{t|t-1}, x_{t|t}, P_{t|t-1}$ et $P_{t|t}$ $t = 1, \dots, T$; ainsi que $x_{t|T}$ et $P_{t|T}$. Par la suite on génère $x_t^{(i)}$ de la fonction densité $\frac{1}{3}N(x_{t|t-1}, cPt|t-1) + \frac{1}{3}N(x_{t|t}, cPt|t) + \frac{1}{3}N(x_{t|T}, cPt|T)$, où $c = 4$.

Pour chaque $t = 1, \dots, T$ on ajuste $p(x_t^{(i)} | \mathbf{y}_T)$ ($i = 0, 1, \dots, M$) en la multipliant par une constante telle que:

$$\frac{1}{2} \sum_{i=1}^M (x_t^{(i)} - x_t^{(i-1)}) [p(x_t^{(i-1)} | \mathbf{y}_T) + p(x_t^{(i)} | \mathbf{y}_T)] = 1$$

Le tableau suivant montre la moyenne et le RMSE (*Root Mean Squared Error*) de différents estimateurs. Les résultats du NFML, GMM, SML et MCMC sont respectivement obtenus de la table I de Watanabe (1999), de la table 5 et de la table 9 de Andersen et Sorensen (1996), table 2 de Danielsson (1994b) et de la table 9 de Jacquier (1994). Voir les références pour une définitions précises de ces estimateurs.

Méthode	α	β	σ_ω	α	β	σ_ω	α	β	σ_ω
	-0.736	0.900	0.363	-0.368	0.950	0.260	-0.147	0.980	0.166
NFML									
N=25	-0.812 (0.199)	0.890 (0.027)	0.406 (0.068)	-0.426 (0.124)	0.942 (0.017)	0.294 (0.052)	-0.194 (0.083)	0.974 (0.011)	0.197 (0.043)
N=50	-0.776 (0.168)	0.895 (0.023)	0.368 (0.041)	-0.406 (0.106)	0.945 (0.014)	0.264 (0.032)	-0.178 (0.067)	0.976 (0.009)	0.169 (0.024)
QML	-0.837 (0.369)	0.887 (0.050)	0.383 (0.095)	-0.419 (0.187)	0.943 (0.025)	0.271 (0.061)	-0.182 (0.087)	0.975 (0.012)	0.174 (0.037)
GMM	-0.592 (0.311)	0.920 (0.042)	0.279 (0.115)	-0.286 (0.197)	0.961 (0.027)	0.190 (0.099)	-0.140 (0.112)	0.981 (0.015)	0.125 (0.068)
SML	-0.721 (0.15)	0.902 (0.02)	0.359 (0.039)						
MCMC	-0.762 (0.15)	0.896 (0.02)	0.359 (0.034)						

Tableau 6.1: L'estimateur de la moyenne selon Watanabe et le RMSE(Root mean squared error) comparé à d'autres estimateurs [Watanabe, 1999]

6.4 Théorie de filtrage continu-discret non linéaire

Pour être capable de prédire l'évolution de X_t dans (6.3) nous avons besoin d'informations sur X_t et sur son comportement.

Il est connu que le processus \mathbf{X}_t généré par l'équation (6.3) est un processus markovien (voir (Jazwinski, 1970, p.105-110) ou (Karatzas et Shreve, 1996, p.74-75)). Sachant que X_t est un processus markovien, il en découle que X_t est totalement caractérisé par la densité de probabilité $p_{x_t}(x_t, t)$ et la fonction de probabilité de transition conditionnelle $p_{x_t|y_{t_{k-1}}}(x_t, t|y_{t_{k-1}})$ ⁶.

Le problème consiste à calculer la densité conditionnelle $p_{x_{t_k}|y_{t_k}}(x_{t_k}, t_k|y_{t_k})$ de

⁶ $\mathcal{Y}_{t_{k-1}} = \sigma\{y_{t_1}, \dots, y_{t_{k-1}}\}$

X_t , sachant les observations faites jusqu'à l'instant t_k , ; ce que nous pouvons accomplir de la manière suivante: Premièrement, on considère la propagation de $p_{x_t|\mathcal{Y}_{t_{k-1}}}(x_t, t|\mathcal{Y}_{t_{k-1}})$ entre les temps d'échantillonnage t_{k-1} et t_k . Comme on sait que la solution de l'équation (6.3) est un processus markovien et par conséquent caractérisé par $p_x(x, t|\zeta, t')$ où $t \geq t'$; si on connaît $p_x(x_t, t|\zeta, t_{k-1})$ pour chaque $t \in [t_{k-1}, t_k]$, alors la densité conditionnelle $p_{x_t|\mathcal{Y}_{t_{k-1}}}(x_t, t|\mathcal{Y}_{t_{k-1}})$ peut être établie pour chaque t dans cet intervalle, et donc peut être propagée sur tout l'intervalle d'échantillonnage par l'équation suivante:

$$p_{x_t|\mathcal{Y}_{t_{k-1}}}(x_t, t|\mathcal{Y}_{t_{k-1}}) = \int_{-\infty}^{+\infty} p_x(x_t, t|\zeta, t_{k-1}) p_{x_{t_{k-1}}|\mathcal{Y}_{t_{k-1}}}(\zeta, t_{k-1}|\mathcal{Y}_{t_{k-1}}) d\zeta \quad (6.26)$$

et ce en utilisant l'équation de Chapman-Kolmogorov et en calculant $p_{x_{t_{k-1}}|\mathcal{Y}_{t_{k-1}}}(\zeta, t_{k-1}|\mathcal{Y}_{t_{k-1}})$ des mesures précédentes. Pour résoudre l'équation (6.26) nous devons calculer la quantité $p_x(x_t, t|\zeta, t_{k-1})$ ce qui nous amène à l'équation avant de Kolmogorov:

Équation de rétro-propagation de Kolmogorov(ERPK) L'évolution de la densité de transition du processus de Markov généré par l'équation (6.3) $p_x(x_t, t|\zeta, t')$ où $t \geq t'$, peut être décrit dans le cas scalaire par l'EDP:

$$\frac{\partial}{\partial t} p_x(x, t|\zeta, t') = -\frac{\partial \{p_x(x, t|\zeta, t') f(x, t)\}}{\partial x} + \frac{1}{2} \frac{\partial^2 \{p_x(x, t|\zeta, t') g^2(x, t)\}}{\partial x^2}, \quad t > t'. \quad (6.27)$$

Voir [Jazwinski, 1970, p. 126-129]). Cette équation est obtenue en utilisant l'équation de Chapman-Kolmogorov dans une expansion de Taylor de $p_x(x, t|\zeta, t')$ (voir [Maybeck, 1982, p.193-196]).

On est capable maintenant, grâce à l'ERPK, d'évaluer $p_x(x_t, t|\zeta, t_{k-1})$ pour tout $t \in [t_{k-1}, t_k]$, ce qui nous permet de calculer $p_{x_{t_k}|\mathcal{Y}_{t_{k-1}}}(x_{t_k}, t_k|\mathcal{Y}_{t_{k-1}})$ à partir de $p_{x_{t_{k-1}}|\mathcal{Y}_{t_{k-1}}}(x_{t_{k-1}}, t_{k-1}|\mathcal{Y}_{t_{k-1}})$ en utilisant l'équation (6.26). On peut montrer que $p_{x_{t_k}|\mathcal{Y}_{t_{k-1}}}(x_{t_k}, t_k|\mathcal{Y}_{t_{k-1}})$ satisfait l'ERPK dans l'intervalle $t \in [t_{k-1}, t_k]$, partant de la densité $p_{x_{t_{k-1}}|\mathcal{Y}_{t_{k-1}}}(x_{t_{k-1}}, t_{k-1}|\mathcal{Y}_{t_{k-1}})$ obtenu à partir de la mise à jour de la mesure précédente. Nous voudrions maintenant incorporer la mesure $Y_{t_k} = \xi_k$ qui est observable à l'instant t_k . En appliquant la formule de Bayes, on établit une estimée de la mesure mis à jour:

$$p_{x_{t_k}|\mathcal{Y}_{t_k}}(x_{t_k}, t_k|\mathcal{Y}_{t_k}) = \frac{p_{y_{t_k}|x_{t_k}, \mathcal{Y}_{t_{k-1}}}(\xi_k|x_{t_k}, \mathcal{Y}_{t_{k-1}})p_{x_{t_{k-1}}|\mathcal{Y}_{t_{k-1}}}(x_{t_{k-1}}, t_{k-1}|\mathcal{Y}_{t_{k-1}})}{p_{y_{t_k}|\mathcal{Y}_{t_{k-1}}}(\xi_k|\mathcal{Y}_{t_{k-1}})} \quad (6.28)$$

ce qui peut être utilisé comme nouvelle valeur initiale dans l'ERPK.

Conceptuellement, cette méthode fournit un outil analytique pour résoudre le problème de filtrage, mais les insuffisances computationnelles actuelles rendent l'implémentation d'une telle solution inappropriée. Dès lors, il devient nécessaire d'introduire quelques approximations sous la forme d'hypothèses simplificatrices.

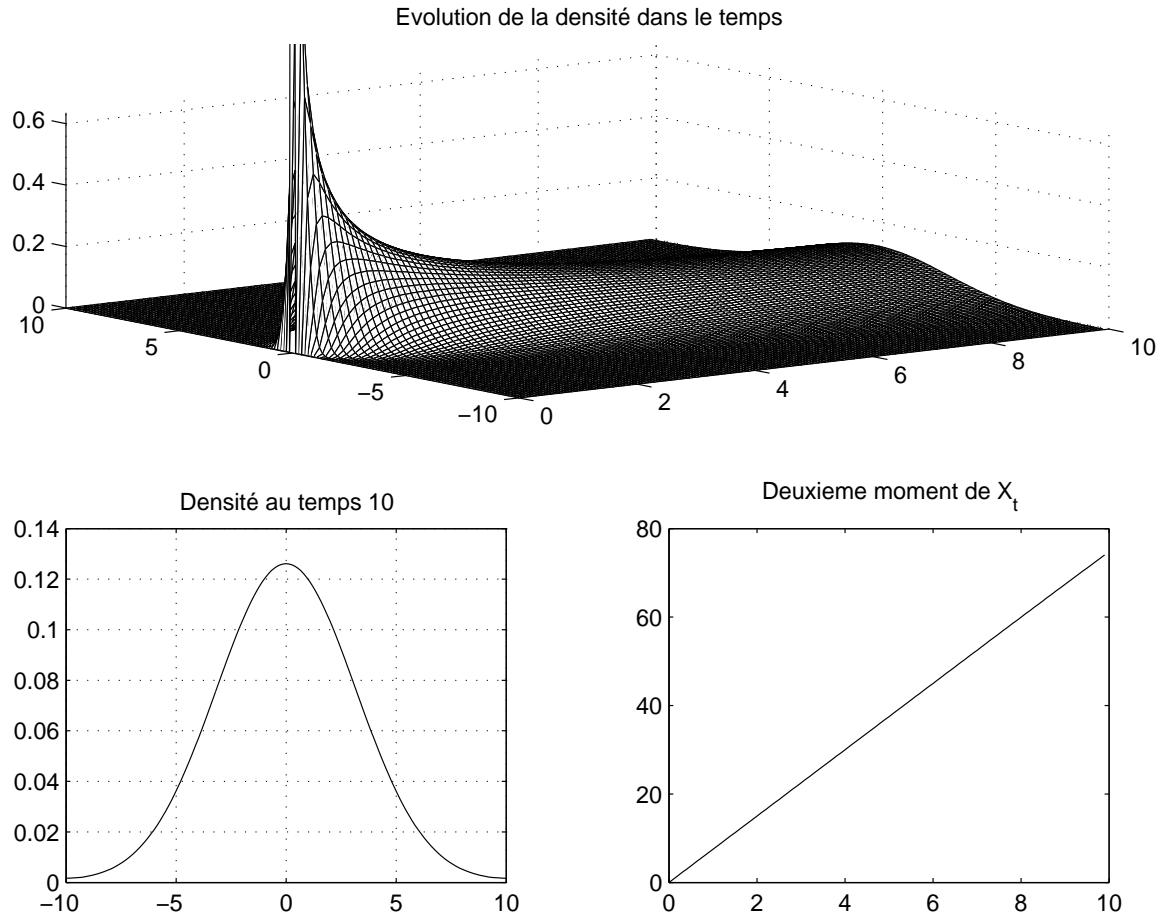


Figure 6.1: Solution de l'équation avant de Kolmogorov dans le cas d'un processus de Wiener

6.5 Estimateur des moments conditionnels

Dans le précédent paragraphe, nous avons présenté une solution du problème de filtrage pour les distributions de transition conditionnelles d'un état du système. La prédiction d'un tel problème passait par la solution de l'équation

de rétro-propagation de Kolmogorov(ERPK) et en utilisant l'équation des observations et des règles de Bayes pour une mise à jour de la distribution après chaque nouvelle mesure. Cette méthode exige de lourds calculs, ce qui nous pousse à réduire la distribution à une quantité plus facilement manipulable. Une approche commune en théorie du filtrage, consiste à considérer la moyenne conditionnelle et la covariance de l'état, respectivement notées $\hat{x}_{t|t_{k-1}}$ et $P_{t|t_{k-1}}$.

Considérons la propagation temporelle de ces deux moments:

$$\hat{x}_{t|t_{k-1}} = E[x_t | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]$$

$$P_{t|t_{k-1}} = E[(x_t - \hat{x}_{t|t_{k-1}})(x_t - \hat{x}_{t|t_{k-1}})^\top | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]$$

Pour une certaine réalisation du système $\mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}$.

Entre deux observations faites respectivement aux instants t_{k-1} et t_k , la densité conditionnelle $p_{x_{t_k}|\mathcal{Y}_{t_{k-1}}}(x_{t_k}, t_k | \mathcal{Y}_{t_{k-1}})$ et la densité $p_{x_t}(x_t, t)$ satisfont l'ERPK; donc les propagations de $\hat{x}_{t|t_{k-1}}$ et $P_{t|t_{k-1}}$ évoluent tels que

$$\hat{x}_{t|t_{k-1}} = E[f(x_t, t) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]$$

et

$$\begin{aligned} P_{t|t_{k-1}} &= E[f(x_t, t)x_t^\top | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}] - E[f(x_t, t) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]x_{t|t_{k-1}}^\top \\ &+ E[x_t f_T(x_t, t) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}] - x_{t|t_{k-1}}^\top E[f_T(x_t, t) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}] \\ &+ E[g(x_t, t)Q_t q^\top(x_t, t) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]. \end{aligned}$$

Pour la démonstration, voir [Maybeck, 1982, p.199-201].

Ces moments sont alors propagés jusqu'à l'instant t_k pour produire $\hat{x}(t_k^-) = \hat{x}_{t_k|t_{k-1}}$ et $P(t_k^-) = P_{t_k|t_{k-1}}$. Pour trouver l'estimée mise à jour au temps t_k , nous devons incorporer les mesures valables à l'instant t_k dans le calcul de la moyenne conditionnelle.

$$\hat{x}(t_k^+) = E[x_{t_k} | \mathcal{Y}_{t_k} = y_k] = \int_{-\infty}^{\infty} \zeta p_{x_{t_k} | \mathcal{Y}_{t_k}}(\zeta, t_k | \mathcal{Y}_{t_k}) d\zeta.$$

où $\hat{x}(t_k^+) = \hat{x}_{t_k|t_k}$ et $p_{x_{t_k} | \mathcal{Y}_{t_k}}(\zeta, t_k | \mathcal{Y}_{t_k})$ est obtenu à partir de la formule de Bayes. Cette estimée peut être exprimée de manière équivalente

$$\hat{x}(t_k^+) = \frac{E[x_{t_k} p_{y_{t_k} | x_{t_k}}(y_k | x_{t_k}) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]}{E[p_{y_{t_k} | x_{t_k}}(y_k | x_{t_k}) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]}.$$

Et de manière similaire,

$$\begin{aligned} P(t_k^+) &= \int_{-\infty}^{\infty} \zeta \zeta^\top p_{x_{t_k} | \mathcal{Y}_{t_k}}(\zeta, t_k | \mathcal{Y}_{t_k}) d\zeta - \hat{x}(t_k^+) \hat{x}^\top(t_k^+) \\ &= \frac{E[x_{t_k} x_{t_k}^\top p_{y_{t_k} | x_{t_k}}(y_k | x_{t_k}) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]}{E[p_{y_{t_k} | x_{t_k}}(y_k | x_{t_k}) | \mathcal{Y}_{t_{k-1}} = \mathbf{y}_{k-1}]} - \hat{x}(t_k^+) \hat{x}^\top(t_k^+) \end{aligned}$$

Les quantités conditionnelles dans les équations précédentes requièrent la connaissance de la densité $p_{x_{t_k} | \mathcal{Y}_{t_{k-1}}}(x_k, t_k | \mathcal{Y}_{t_{k-1}})$, et donc en faisant certaines suppositions sur cette densité, nous serons capable d'obtenir des algorithmes réalisables d'un point de vue computationnel. La section suivant décrit deux filtres qui se basent sur des suppositions simplificatrices sur la fonction de densité conditionnelle.

6.6 Filtres non linéaires approximatifs

Pour développer des algorithmes réalisables pour la prédiction de \hat{x}_t et P_t , il est nécessaire de faire des hypothèses sur les densités de probabilité conditionnelle. On peut supposer par exemple que la densité est quasi-symétrique, donc les moments d'ordre 3 et les moments d'ordres impairs supérieurs sont approximativement égaux à 0, et que cette fonction est suffisamment concentré autour de la moyenne, pour supposer que les moments d'ordre 4 et les moments d'ordre supérieur sont assez petits pour être négligeables. Ces différentes suppositions nous amènent au filtre du second ordre tronqué(TSOF). Une autre alternative est de supposer que la densité conditionnelle est quasi-gaussienne, donc les moments d'ordre 3 et les moments d'ordre impairs supérieurs sont égaux à zéro, et que les moments d'ordre 4 et les moments supérieurs peuvent être exprimés en terme de covariance. Ces suppositions nous amènent au filtre gaussien du second ordre (GSOF). Nous présentons dans cette section, ces deux filtres dans leurs version scalaire.

6.6.1 Version scalaire du filtre tronqué du second ordre (TSOF)

On considère la version scalaire du système d'équation (6.3) et (6.4). On voudrait pouvoir prédire l'évolution de la moyenne conditionnelle \hat{x}_t et sa variance conditionnelle P_t , sans connaissance préalable de la densité conditionnelle $p_x(x_t, t | \zeta, t_{k-1})$ pour X_t . Pour alléger la notation, nous noterons $p_x(x_t, t | \zeta, t_{k-1})$ dans ce paragraphe par $p_x(x_t, t)$.

Équations de propagation temporelle pour la version scalaire du TSOF:

On considère que $\varphi(X_t)$ est une fonction de X_t . L'espérance de $\varphi(X_t)$ est par définition:

$$E[\varphi(X_t)] = \int \varphi(x_t) p_x(x_t, t) dx_t$$

et donc

$$\frac{dE[\varphi(X_t)]}{dt} = \int \varphi(x_t) \frac{\partial p_x(x_t, t)}{\partial t} dx_t \quad (6.29)$$

En utilisant l'ERPK (6.27), l'équation (6.29) devient

$$\frac{dE[\varphi(X_t)]}{dt} = - \int \varphi(x_t) \frac{\partial(p_x(x_t, t) f)}{\partial x_t} dx_t + \frac{1}{2} Q_t \int \varphi(x_t) \frac{\partial^2(p_x(x_t, t) g^2)}{\partial x_t^2} dx_t.$$

En faisant une intégration par parties, sous l'hypothèse que $p_x(x_t, t)$ est bien défini à l'infini, on a

$$\begin{aligned} \frac{dE[\varphi(X_t)]}{dt} &= -p_x(x_t, t) f \varphi(x_t) + \int p_x(x_t, t) f \frac{\partial \varphi(x_t)}{\partial x_t} dx_t \\ &+ \frac{1}{2} Q_t \frac{\partial(p_x(x_t, t) g^2)}{\partial x_t} \varphi(x_t) dx_t \\ &+ \frac{1}{2} Q_t \int \frac{\partial(p_x(x_t, t) g^2)}{\partial x_t} \frac{\partial \varphi(x_t)}{\partial x_t} dx_t \end{aligned} \quad (6.30)$$

L'hypothèse faite sur $p_x(x_t, t)$, provient du fait que $p_x(x_t, t)$ est une densité de probabilité, et donc $\int_{-\infty}^{\infty} p_x(x_t, t) dt = 1$, d'où

$$\lim_{x \rightarrow \infty} p_x(x_t, t) = \lim_{x \rightarrow -\infty} p_x(x_t, t) = 0,$$

et par conséquent

$$\lim_{x \rightarrow \infty} p_x(x_t, t) f \varphi(x_t) = \lim_{x \rightarrow -\infty} p_x(x_t, t) f \varphi(x_t) = 0$$

$$\lim_{x \rightarrow \infty} \frac{\partial(p_x(x_t, t)g^2)}{\partial x_t} \varphi(x_t) = \lim_{x \rightarrow -\infty} \frac{\partial(p_x(x_t, t)g^2)}{\partial x_t} \varphi(x_t) = 0.$$

En utilisant une seconde fois une intégration par partie sur la dernière intégrale de (6.30), on obtient l'expression suivante

$$\begin{aligned} \frac{dE[\varphi(X_t)]}{dt} &= \int p_x(x_t, t) f \frac{\partial \varphi(x_t)}{\partial x_t} dx_t \\ &- \frac{1}{2} Q_t \left\{ \left\{ p_x(x_t, t) g^2 \frac{\partial \varphi(x_t)}{\partial x_t} \right\} - \int p_x(x_t, t) g^2 \frac{\partial^2 \varphi(x_t)}{\partial x_t^2} dx_t \right\} \\ &= E \left[f \frac{\partial \varphi(x_t)}{\partial x_t} \right] + \frac{1}{2} Q_t E \left[g^2 \frac{\partial^2 \varphi(x_t)}{\partial x_t^2} \right]. \end{aligned} \tag{6.31}$$

En posant $\varphi(x) = x$ et en calculant les dérivées partielles de l'équation (6.31), on obtient:

$$\frac{d\hat{X}_t}{dt} = \hat{f}(x_t, t) \tag{6.32}$$

et en posant $\varphi(x) = x^2$, on obtient:

$$\frac{d\hat{X}_t^2}{dt} = 2\hat{x}_t f(x_t, t) + Q_t g^2(x_t, t)$$

tels que: $\hat{x}_t f(x_t, t) = E[x_t f(x_t, t)]$ et $\hat{g}^2(x_t, t) = E[g^2(x_t, t)]$ or

$$\frac{d\hat{x}_t^2}{dt} = 2\hat{x}_t \frac{d\hat{x}_t}{dt}$$

et donc

$$d\hat{x}_t^2 = 2\hat{x}_t \hat{f}(x_t, t) dt.$$

Sachant que $V[X_t] = E[X_t^2] - E[X_t]^2$, on peut obtenir une expression équivalente

pour la variance conditionnelle:

$$\begin{aligned}\frac{dP_t}{dt} &= \frac{dE[x_t^2]}{dt} - \frac{d\hat{x}_t^2}{dt} \\ &= 2\hat{x}_t f(x_t, t) - \hat{x}_t \hat{f}(x_t, t) + Q_t \hat{g}^2(x_t, t)\end{aligned}$$

En résumé, \hat{x}_t et P_t satisfont:

$$\begin{cases} \frac{d\hat{x}_t}{dt} = \hat{f}(x_t, t) \\ \frac{dP_t}{dt} = 2\hat{x}_t f(x_t, t) - \hat{x}_t \hat{f}(x_t, t) + Q_t \hat{g}^2(x_t, t) \end{cases}$$

Et comme $\frac{d\hat{x}_t}{dt} = \int f(x_t, t) p_x(x_t, t) dx_t$, nous avons besoin d'avoir une connaissance sur la densité conditionnelle $p_x(x_t, t)$, mais comme cette information n'est pas disponible, il est possible simplement d'obtenir des équations différentielles approximatives en appliquant les hypothèses expliquées auparavant à $p_x(x_t, t)$. En utilisant la formule de Taylor, on obtient:

$$\frac{d\hat{x}_t}{dt} = f(\hat{x}_t, t) + f_x(\hat{x}_t, t)(x_t - \hat{x}_t) + \frac{1}{2} f_{xx}(\hat{x}_t, t)(x_t - \hat{x}_t)^2 + \dots$$

où $f_{xx}(\hat{x}_t, t) = \frac{\partial^2 f(\hat{x}_t, t)}{\partial x_t^2}$ et donc

$$\frac{d\hat{x}_t}{dt} = f(\hat{x}_t, t) + \frac{1}{2} P_t f_{xx}(\hat{x}_t, t)$$

De la même manière, en développant g^2 on obtient une approximation de $\frac{dP_t}{dt}$:

$$\frac{dP_t}{dt} = 2f_x(\hat{x}_t, t)P_t + Q_t g^2(\hat{x}_t, t) + P_t g_x^2(\hat{x}_t, t) + P_t g(\hat{x}_t, t)g_{xx}(\hat{x}_t, t)$$

En résumé, l'évolution temporelle de \hat{x}_t et P_t est décrite par le système d'équations différentielles suivant:

$$\begin{aligned}
\frac{d\hat{x}_{t|t_{k-1}}}{dt} &= f(\hat{x}_{t|t_{k-1}}, t) + \frac{1}{2} P_{t|t_{k-1}} f_{xx}(\hat{x}_{t|t_{k-1}}, t) \\
\frac{dP_{t|t_{k-1}}}{dt} &= 2f_x(\hat{x}_{t|t_{k-1}}, t)P_{t|t_{k-1}} + Q_t g^2(\hat{x}_{t|t_{k-1}}, t) + P_{t|t_{k-1}} g_x^2(\hat{x}_{t|t_{k-1}}, t) \\
&\quad + P_{t|t_{k-1}} g(\hat{x}_{t|t_{k-1}}, t)g_{xx}(\hat{x}_{t|t_{k-1}}, t)
\end{aligned}$$

Ce système est dit système d'équation de propagation du filtre tronqué du second ordre.

Équations de mise à jour du TSOF:

Les équations de mise à jour du TSOF sont inspiré par celle du filtre de Kalman ordinaire, qui assument que ces valeurs mises à jour; $\hat{x}_{t_k^+}$ et $P_{t_k^+}$ ont la structure suivante:

$$\begin{cases} \hat{x}_{t_k^+} = a_0 + a_1 \left\{ Y_{t_k} - \hat{Y}_{t_K^-} \right\} \\ P_{t_k^+} = b_0 + b_1 \left\{ Y_{t_k} - \hat{Y}_{t_K^-} \right\} \end{cases}$$

Il nous faut déterminer les constantes a_0, a_1, b_0 et b_1 sous l'hypothèse que $P_{t_k^+}$ doit être positive. Cette détermination peut être faite en faisant appel à des fonctions arbitraires des variables aléatoires X_{t_k} et Y_{t_k} , respectivement $\psi(\cdot)$ et $\theta(\cdot)$. Nous renvoyons le lecteur à [Maybeck, 1982, p.219-220]. Les

constantes sont données par:

$$\begin{cases} a_0 = \hat{x}_{t_k^-} \\ a_1 = \left(\hat{x}\hat{h} - \hat{x}\hat{h} \right) \left\{ (\hat{h} - \hat{h})^2 + Vt_k \right\}^{-1} \\ b_0 = P_{t_k^-} - a_1 \left(\hat{x}\hat{h} - \hat{x}\hat{h} \right) \\ b_1 = \left\{ (\hat{h} - \hat{h})^2 h - 2a_1 \left\{ (\hat{h} - \hat{h})(\hat{x} - \hat{x})h \right\} + a_1^2 \left\{ (\hat{h} - \hat{h})^2 h + R\hat{h} \right\} - b_0 \hat{h} \right\} \\ \times \left\{ (\hat{h} - \hat{h})^2 + Vt_k \right\}^{-1} \end{cases}$$

Ce qui donne en utilisant un développement en série de Taylor tronqué au second ordre:

$$\begin{aligned} \hat{y}_{t_k^-} &= h(\widehat{x_{t_k}}, \widehat{t_k}) \cong (\hat{x}_{t_k^-}, t_k) + \frac{1}{2} h_{xx}(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \\ \left(\widehat{x}\hat{h} - \hat{x}\hat{h} \right) &\cong h_x(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \\ (\widehat{h} - \hat{h})^2 &\cong h_x^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-} - \frac{1}{4} h_{xx}^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-}^2 \\ b_1 &\cong -\frac{1}{2} P_{t_k^-}^2 h_{xx}(\hat{x}_{t_k^-}, t_k) \left\{ (\widehat{h} - \hat{h})^2 + Vt_k \right\}^{-1} \\ &\times \left\{ 1 - 3P_{t_k^-} h_{xx}(\hat{x}_{t_k^-}, t_k) \left\{ (\widehat{h} - \hat{h})^2 + Vt_k \right\}^{-1} \right. \\ &\quad \left. - P_{t_k^-} h_x^2(\hat{x}_{t_k^-}, t_k) \left\{ (\widehat{h} - \hat{h})^2 + Vt_k \right\}^{-2} \right. \\ &\quad \left. \times \left\{ \frac{1}{4} P_{t_k^-}^2 h_{xx}^2(\hat{x}_{t_k^-}, t_k) + Vt_k \right\} \right. \\ &\quad \left. + 2P_{t_k^-}^2 h_{xx}^4(\hat{x}_{t_k^-}, t_k) \left\{ (\widehat{h} - \hat{h})^2 + Vt_k \right\}^{-2} \right\} \end{aligned}$$

En combinant ces résultats, avec ceux de la construction des équations de mises à jour obtenues plus haut on obtient les équations de mises à jour du filtre tronqué du second ordre:

$$\begin{aligned}
A_{t_k} &= h_x^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-} - \frac{1}{4} h_{xx}^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-}^2 + V t_k & (6.33) \\
K_{t_k} &= P_{t_k^-} h_x(\hat{x}_{t_k^-}, t_k) A_{t_k}^{-1} \\
\hat{x}_{t_k^+} &= \hat{x}_{t_k^-} + K_{t_k} \left\{ y_k - h(\hat{x}_{t_k^-}, t_k) - \frac{1}{2} h_{xx}(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \right\} \\
P_{t_k^+} &= P_{t_k^-} - K_{t_k} h_x(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \\
&\quad + b_1 \left\{ y_k - h(\hat{x}_{t_k^-}, t_k) - \frac{1}{2} h_{xx}(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \right\}
\end{aligned}$$

6.6.2 Version scalaire du filtre gaussien du second ordre (GSOF)

Tandis que le filtre TSOF a ignoré tous les moments centraux de X_t au-delà du second ordre, le filtre gaussien du second ordre (GSOF) prend en considération les moments centraux d'ordre 3 et 4 dans l'approximation des équations de propagation.

Équations de propagation temporelle pour la version scalaire du GSOF:

Comme nous l'avons expliqué plus haut, le GSOF diffère du TSOF par l'approximation du moment central d'ordre 4 de $p_x(x_t, t)$ avec la valeur d'hypothèse que $p_x(x_t, t)$ soit gaussienne. Le moment central d'ordre 4 est donné par:

$$\widehat{(x - \hat{x})^4} = 3P^2$$

où P est la matrice des covariances conditionnelles. Et avec le même raisonnement que dans le cas du TSOF, nous obtenons les équations de propagation du GSOF:

$$\begin{cases} \frac{d\hat{x}_{t|t_{k-1}}}{dt} = f(\hat{x}_{t|t_{k-1}}, t) + \frac{1}{2}P_{t|t_{k-1}}f_{xx}(\hat{x}_{t|t_{k-1}}, t) \\ \frac{dP_{t|t_{k-1}}}{dt} = 2f_x(\hat{x}_{t|t_{k-1}}, t)P_{t|t_{k-1}} + Q_t g^2(\hat{x}_{t|t_{k-1}}, t) + P_{t|t_{k-1}}g_x^2(\hat{x}_{t|t_{k-1}}, t) \\ + P_{t|t_{k-1}}g(\hat{x}_{t|t_{k-1}}, t)g_{xx}(\hat{x}_{t|t_{k-1}}, t) \\ + \frac{3}{4}P_{t|t_{k-1}}^2g_{xx}^2(\hat{x}_{t|t_{k-1}}, t) \end{cases}$$

On notera que les deux filtres diffèrent seulement lorsque g est non linéaire en x_t .

Équations de mise à jour du GSOF:

Pour trouver les équations de mise à jour du filtre gaussien du second ordre, nous assumons la même construction de $\hat{x}_{t_k^+}$ et $P_{t_k^+}$ que pour le TSOF et les constantes a_0, a_1, b_0 et b_1 sont égales à ceux du TSOF. La seule différence entre les deux filtres est le calcul de $\widehat{(h - \hat{h})^2}$ et b_1 qui sont dans ce cas calculés de la manière suivante:

$$\begin{aligned} \widehat{(h - \hat{h})^2} &\cong h_x^2(\hat{x}_{t_k^-}, t_k)P_{t_k^-} - \frac{1}{2}h_{xx}^2(\hat{x}_{t_k^-}, t_k)P_{t_k^-}^2 \\ b_1 &\cong P_{t_k^-}^2 h_{xx}(\hat{x}_{t_k^-}, t_k) \left\{ \widehat{(h - \hat{h})^2} + Vt_k \right\}^{-1} \\ &\quad \times \left\{ 1 - \frac{9}{2}P_{t_k^-}h_{xx}(\hat{x}_{t_k^-}, t_k) \left\{ \widehat{(h - \hat{h})^2} + Vt_k \right\}^{-1} \right. \\ &\quad \left. + \frac{1}{2}P_{t_k^-}h_x^2(\hat{x}_{t_k^-}, t_k) \left\{ \widehat{(h - \hat{h})^2} + Vt_k \right\}^{-2} \right. \\ &\quad \left. \times \left\{ 7P_{t_k^-}h_{xx}^2(\hat{x}_{t_k^-}, t_k) + Vt_k \right\} \right\} \end{aligned}$$

Les équations de mise à jour du filtre gaussien du second ordre :

$$\begin{aligned}
 A_{t_k^-} &= h_x^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-} - \frac{1}{2} h_{xx}^2(\hat{x}_{t_k^-}, t_k) P_{t_k^-}^2 + V t_k \\
 K_{t_k^-} &= P_{t_k^-} h_x(\hat{x}_{t_k^-}, t_k) A_{t_k^-}^{-1} \\
 \hat{x}_{t_k^+} &= \hat{x}_{t_k^-} + K_{t_k^-} \left\{ y_k - h(\hat{x}_{t_k^-}, t_k) - \frac{1}{2} h_{xx}(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \right\} \\
 P_{t_k^+} &= P_{t_k^-} - K_{t_k^-} h_x(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \\
 &\quad + b_1 \left\{ y_k - h(\hat{x}_{t_k^-}, t_k) - \frac{1}{2} h_{xx}(\hat{x}_{t_k^-}, t_k) P_{t_k^-} \right\}
 \end{aligned}$$

Chapitre 7

Résultats et conclusions

Dans le présent chapitre, le filtre gaussien du second ordre sera appliqué au modèle de Black-Scholes-Courtadon, aussi nous calculerons la fonction du maximum de vraisemblance. Nous appliquerons par la suite le filtre aux données financières des quatres séries que nous avons présenté au chapitre 2. Nous conclurons le chapitre par une discussion des résultats obtenus ainsi que des suggestions pour les études futures.

7.1 Discrétisation des équations différentielles stochastiques

Dans cette section nous présentons la méthode d'Euler utilisée pour la discrétisation des équations différentielles stochastiques. Nous ne présenterons pas de

preuve des résultats présentés mais seulement quelques exemples.

7.1.1 Méthode d'Euler

On considère le modèle suivant

$$X_t = X_0 + \sum_{j=1}^d \int_0^t a^j(s, X_s) ds + \sum_{j=1}^m \int_0^t b^j(s, X_s) dW_s^j$$

où m est la dimension du processus de Wiener. La méthode d'Euler, ou méthode d'Euler-Maruyama, est caractérisée par l'utilisation des deux premiers termes dans l'expansion de Taylor de l'EDS. Pour le cas unidimensionnel, où $m = d = 1$ le schéma d'Euler prend la forme suivante:

$$Y_{n+1} = Y_n + a\Delta + b\Delta W$$

où $\Delta = \tau_{n+1} - \tau_n$ est la longueur de la discrétisation temporelle, et $\Delta W_{n+1} = W_{\tau_{n+1}} - W_{\tau_n}$ est l'incrément de loi $N(0, \Delta)$ du processus de Wiener W sur l'intervalle $[\tau_{n+1}, \tau_n]$.

Exemple: Soit le modèle univarié suivant

$$dS_t = \alpha S_t dt + \sigma S_t^\gamma dW_t,$$

où α et σ sont deux constantes positives. La discrétisation de cette EDS selon la méthode d'Euler donne

$$Y_{n+1} = Y_n + \alpha Y_n \Delta + \sigma Y_n^\gamma \Delta W_{n+1}$$

où $\Delta W \sim N(0, \Delta)$. Il faut par ailleurs s'assurer que $Y_n > 0, \forall n$

On considère maintenant le cas multidimensionnel. Dans ce cas, le schéma d'Euler est donné par:

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W_{n+1}^j \quad (7.1)$$

où k réfère au k -ième composant du vecteur discrétisé. Le facteur $b^{k,j}$ réfère au composant (k, j) de la matrice b .

Dans le cas d'un modèle corrélé bivarié, donné par le système suivant

$$\begin{aligned} dS_t &= \alpha S_t dt + \sigma S_t^\gamma dW_{1,t} \\ d\sigma_t &= \kappa(\beta - \sigma_t) dt + \xi \sigma_t dW_{2,t} \end{aligned}$$

où $W_{1,t}$ et $W_{2,t}$ sont deux processus de Wiener corrélés avec un coefficient de corrélation ρ . Il est à remarquer que dans le cas où $\theta = 0$, on retrouve le modèle de Black-Scholes; et dans ce cas, la volatilité σ_t est déterministe et converge vers β . On peut introduire une transformation sur nos deux processus de Wiener pour éliminer le coefficient de corrélation. On obtient le système d'équations suivant:

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t^\gamma & 0 \\ \xi \rho \sigma_t & \xi \sqrt{1 - \rho^2} \sigma_t \end{pmatrix} \begin{pmatrix} d\tilde{W}_{1,t} \\ d\tilde{W}_{2,t} \end{pmatrix}$$

où $\tilde{W}_{1,t}$ et $\tilde{W}_{2,t}$ sont deux processus de Wiener non corrélés. En utilisant l'équation (7.1), le schéma d'Euler pour le système corrélé bivarié est donné par:

$$\begin{pmatrix} Y_{n+1}^1 \\ Y_{n+1}^2 \end{pmatrix} = \begin{pmatrix} Y_n^1 \\ Y_n^2 \end{pmatrix} + \begin{pmatrix} \alpha Y_n^1 \\ \kappa(\beta - Y_n^2) \end{pmatrix} \Delta + \begin{pmatrix} Y_n^2 (Y_n^1)^\gamma \Delta W_1 \\ \xi \rho Y_n^2 \Delta W_1 + \xi \sqrt{1 - \rho^2} Y_n^2 \Delta W_2 \end{pmatrix}$$

où $\Delta W_1 \sim N(0, \Delta)$ et $\Delta W_2 \sim N(0, \Delta)$.

7.2 Méthode du maximum de vraisemblance

Dans cette section, nous présentons la méthode du maximum de vraisemblance basée sur l'hypothèse de normalité de l'estimation de l'erreur $\{y_{t_k} - h(\hat{x}_{t_k}, t_k, \theta)\}$. Cette hypothèse peut être par ailleurs testée en utilisant les tests statistiques standards sur les propriétés d'un bruit blanc gaussien. Soit $X_t, t > t_0$ solution de l'équation:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \quad X_0 = x, \quad (7.2)$$

telle que la condition initiale X_0 est supposée être de loi normale ($X_0 \sim N_d(\mu_0, \sigma_0)$). On suppose par la suite que les fonctions $f(X_t, t)$ et $g(X_t, t)$ sont continues et données par le paramètre inconnu $\theta \in \Theta \subseteq R^p$. La fonction de vraisemblance pour le paramètre θ est données par:

$$\begin{aligned} L(\theta, x^N) &= p_x(x^N | \theta) \\ &= \left[\prod_{k=1}^N p_x(x_{t_k} | x_{t_{k-1}}, \theta) \right] p_x(x_{t_0} | \theta) \end{aligned} \quad (7.3)$$

où $x^N = [x_{t_N}, x_{t_{N-1}}, \dots, x_{t_1}, x_{t_0}]^\top$ est le vecteur échantillon de la variable d'état à l'instant $t_k, k = 1, \dots, N$ et $p_x(x_{t_0} | \theta)$ est la densité de la distribution de x_{t_0} . La fonction de densité de la probabilité conditionnelle $p_x(x_{t_k} | x_{t_{k-1}})$ peut être obtenue comme solution de l'équation avant de Kolmogorov, mais comme le plus souvent nous ne disposons pas d'une solution explicite de cette équation et que le processus X_t est observable avec un certain bruit, nous devons utiliser une autre approche, et ce en utilisant les observations $Y_{t_k}, k = 1, \dots, N$. La fonction de vraisemblance est donnée par:

$$L(\theta, y^N | \mathcal{Y}_{t_0}) = p_y(y^N | \theta)$$

$$= \left[\prod_{k=1}^N p_y(y_{t_k} | \mathcal{Y}_{k-1}, \theta) \right] \quad (7.4)$$

où nous avons conditionné par rapport à toutes les observations précédentes et non seulement par rapport à la dernière observation comme dans l'équation (7.4) et ce pour pallier au problème des observations incomplètes. Le problème dans cette approche est que la fonction de densité de la probabilité conditionnelle $p_y(y_{t_k} | \mathcal{Y}_{k-1}, \theta)$ est rarement disponible, cependant nous pouvons considérer la densité de l'estimée de l'erreur conditionnelle:

$$\epsilon_{t_k}(\theta) = Y_{t_k} - h(\hat{x}_{t_k|t_{k-1}}, t_k, \theta)$$

En supposant la normalité de $\epsilon_{t_k}(\theta)$, la fonction de vraisemblance est complètement caractérisée par les deux premiers moments:

$$\begin{aligned} \hat{y}_{t_k|t_{k-1}} &= E[Y_{t_k} | \mathcal{Y}_{k-1}, \theta] = h(\hat{x}_{t_k|t_{k-1}}, \theta) \\ R_{t_k|t_{k-1}} &= V[Y_{t_k} | \mathcal{Y}_{k-1}, \theta], \end{aligned} \quad (7.5)$$

la fonction de vraisemblance conditionnelle est alors donnée par:

$$L(\theta, y^N | y_{t_0}) = \prod_{k=1}^N \left(\frac{1}{2\pi} \right)^{\frac{m}{2}} \det(R_{t_k|t_{k-1}})^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \epsilon_{t_k}^\top R_{t_k|t_{k-1}}^{-1} \epsilon_{t_k} \right)$$

où m est la dimension du vecteur Y . La log-vraisemblance est donc:

$$\log L(\theta, y^N | y_{t_0}) = -\frac{1}{2} \sum_{k=1}^N \left[\log \det(R_{t_k|t_{k-1}}) + \epsilon_{t_k}^\top R_{t_k|t_{k-1}}^{-1} \epsilon_{t_k} \right] + const, \quad (7.6)$$

7.3 Simulation

Cette section présente les résultats des simulations du modèle de Black&Scholes-Courtadon donné par

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t & 0 \\ 0 & \xi \sigma_t \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix},$$

$$Y_{t_k} = S_{t_k} + e_{t_k}, \quad e_{t_k} \in N(0, \sigma_e^2),$$

où $W_{1,t}$ et $W_{2,t}$ sont deux processus de Wiener corrélés avec un coefficient de corrélation ρ .

Les valeurs des paramètres seront fixées à $\alpha = 0.00198$, $\beta = 0.01$, $\kappa = 0.0071$, $\xi = 0.016$, $\sigma_e = 0.0195$, et $\rho = 0.5$, dans le but de comparaison avec les résultats de Nielsen et all (2000). Comme nous l'avons fait dans le premier paragraphe, nous introduisons une transformation sur nos deux processus de Wiener pour éliminer le coefficient de corrélation. On obtient le système d'équations suivant:

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t & 0 \\ \rho \theta \sigma_t & \sqrt{1 - \rho^2} \theta \sigma_t \end{pmatrix} \begin{pmatrix} d\tilde{W}_{1,t} \\ d\tilde{W}_{2,t} \end{pmatrix}$$

$$Y_{t_k} = S_{t_k} + e_{t_k}, \quad e_{t_k} \in N(0, \sigma_e^2),$$

où $\tilde{W}_{1,t}$ et $\tilde{W}_{2,t}$ sont deux processus de Wiener incorrélatés.

Tous les résultats d'estimation ont été obtenus à travers l'implémentation d'un filtre gaussien du second ordre, présenté dans le chapitre précédent, et les paramètres ont été estimés en utilisant la méthode du maximum de vraisemblance présentée auparavant. Pour la simulation, nous avons considéré 100 séquences de réalisations stochastiques indépendantes, dont chacune contient $N \times 100$ observations. Nous avons fixé le pas de discréétisation égal à $\frac{1}{100}$. Nous avons choisi les tailles $N = 125, 250, 500, 1000$ pour la série des observations $Y_{t_k}, k = 1, \dots, N$. Ces tailles correspondent à 6 mois, 1 an, 2 ans et 4 ans respectivement.

Par la suite, nous comparons la moyenne des paramètres estimés avec les paramètres simulés. Nous calculons par la suite les pourcentages des biais relatifs et des erreurs standards relatives. Le biais relatif, en pourcent, est déterminé par la formule

$$B = 100 \times \frac{1}{n} \sum_{i=1}^n \left(\frac{\theta_i}{\theta_0} - 1 \right),$$

tandis que l'erreur standard relative est déterminée par la formule

$$E = 100 \times \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\theta_i}{\theta_0} - 1 \right)^2},$$

où θ_i est une estimation du paramètre θ_0 .

Pour le calcul de la fonction du maximum de vraisemblance, nous nous servions de la fonction *FMINSEARCH* définie dans Matlab, en reparamétrisant les paramètres pour enlever toutes les contraintes. Ces contraintes sont

données par

$$\begin{cases} \beta > 0, \\ \kappa > 0, \\ \xi > 0, \\ \sigma_e > 0, \\ -1 < \rho < 1. \end{cases}$$

La valeur initiale de $\hat{x}_{t_1|t_2}$ est choisie comme étant la valeur de simulation initiale, et la valeur initiale de $P_{t_1|t_2}$ est choisie de manière aléatoire de valeur proche de 1. D'après l'équation (5.26) du chapitre précédent, les $\hat{x}_{t|t_{k-1}}$ et $P_{t|t_{k-1}}$ satisfont respectivement les équations de propagation suivantes pour $t > t_{k-1}$,

$$x_{0|0}; \frac{dx}{dt} = \begin{pmatrix} \alpha x_1 \\ \kappa(\beta - x_2) \end{pmatrix} \quad (7.7)$$

$$P_{0|0}; \frac{dP}{dt} = \begin{pmatrix} 2\alpha P_{11} & P_{12}(\alpha - \kappa) \\ P_{21}(\alpha - \kappa) & -2\kappa P_{22} \end{pmatrix} + \tilde{P}(x, P), \quad (7.8)$$

sous les conditions initiales $x(t_{k-1}) = \hat{x}_{t_{k-1}|t_{k-1}}$ et $P(t_{k-1}) = P_{t_{k-1}|t_{k-1}}$, où les $\tilde{P}_{jk}(x, P)$, pour $k, j = 1, 2$, sont définis par

$$\begin{aligned} \tilde{P}_{11}(x, P) &= (x_1 x_2)^2 + P_{11} x_2^2 + 2x_1 x_2 (P_{12} + P_{21}) + P_{22} x_1^2 \\ &\quad + \frac{1}{4}(P_{12} + P_{21})^2 + \frac{1}{2}(P_{12}^2 + P_{21}^2 + 2P_{11}P_{22}) \\ \tilde{P}_{21}(x, P) &= \tilde{P}_{12}(x, P) = \xi \rho \{x_1 x_2^2 + P_{21} x_2 + P_{22} x_1 + \frac{1}{2}x_2(P_{12} + P_{21})\} \\ \tilde{P}_{22}(x, P) &= \xi^2(x_2^2 + P_{22}). \end{aligned}$$

Les équations de mises à jours sont

$$\begin{aligned}
\hat{x}_{t_k|t_k} &= \begin{pmatrix} x_1^- \\ x_2^- \end{pmatrix} + \frac{1}{P_{11}^- + \sigma_e^2} \begin{pmatrix} P_{11}^- \\ P_{21}^- \end{pmatrix} \{y_1 - x_1^-\} \\
P_{t_k|t_k} &= \begin{pmatrix} P_{11}^- & P_{12}^- \\ P_{21}^- & P_{22}^- \end{pmatrix} - \frac{1}{P_{11}^- + \sigma_e^2} \begin{pmatrix} P_{11}^{-2} & P_{12}^- P_{11}^- \\ P_{21}^- P_{11}^- & P_{12}^- P_{21}^- \end{pmatrix}, \\
&= \frac{\sigma_e^2}{P_{11}^- + \sigma_e^2} \begin{pmatrix} P_{11}^- & P_{12}^- \\ P_{21}^- & P_{22}^- + \frac{P_{11}^- P_{22}^- - P_{12}^- P_{21}^-}{\sigma_e^2} \end{pmatrix}
\end{aligned}$$

où $P^- = P_{t_k|t_{k-1}}$, et $x^- = \hat{x}_{t_k|t_{k-1}}$. Il faut rappeler que l'on doit utiliser une technique pour garder la matrice des covariances défini positive, et ce en utilisant l'inverse de la matrice des covariances.

Cette formule est donnée par

$$P_{t_k|t_k}^{-1} = P_{t_k|t_{k-1}}^{-1} + \begin{pmatrix} \frac{1}{\sigma_e^2} & 0 \\ 0 & 0 \end{pmatrix}.$$

En effet, puisque $P_{t_k|t_k} = P^- - P^- H^\top (H P^- H^\top + \Sigma)^{-1} H P^-$, avec $H = (1 \ 0)$ et $\Sigma = \sigma_e^2$, on a

$$P_{t_k|t_k}^{-1} = P_{t_k|t_{k-1}}^{-1} + H^\top \Sigma^{-1} H = P_{t_k|t_{k-1}}^{-1} + \begin{pmatrix} \frac{1}{\sigma_e^2} & 0 \\ 0 & 0 \end{pmatrix}.$$

Finalement, pour résoudre un système d'équations différentielles de la forme

$$\frac{dz}{dt} = \phi(z), \quad z(0) = z_0,$$

tel que celui définissant précédent $\hat{x}_{t|t_{k-1}}$ et $P_{t|t_{k-1}}$, on se sert d'une approxi-

	N	α	β	κ	ξ	σ_e	ρ
Biais(%)	100	-9.3	510.4	465.9	-29.3	433.6	-12.0
Erreur(%)		192.1	554.0	673.9	164.6	4357.6	49.7
Biais(%)	250	51.4	399.4	546.7	53.3	599.1	-16.8
Erreur(%)		74.6	426.1	815.9	251.6	1607.1	80.3
Biais(%)	500	-97.0	-6.2	477923.6	117.5	57.9	31.5
Erreur(%)		101.4	7.0	533787.1	2142.6	303.5	47.0
Biais(%)	1000	-99.5	-6.4	557387.2	1248.8	-86.4	38.2
Erreur(%)		100.8	7.1	583934.6	5039.9	92.4	60.4

Tableau 7.1: Résultats de l'estimation des paramètres du modèle de Black-Scholes-Courtadon de la volatilité stochastique.

mation du genre

$$z\left(\frac{j}{k}\right) = z\left(\frac{j-1}{k}\right) + \frac{1}{k}\phi\left(z\left(\frac{j-1}{k}\right)\right), \quad j = 1, \dots, k.$$

Dans notre cas, nous avons choisi $k = 50$.

D'après le tableau des résultats des simulations, on remarque que les estimations des paramètres propres à la volatilité, β , κ et σ_e , ne sont pas très satisfaisantes. Ceci est dû au fait que nous ne disposons que d'informations partielles sur le processus bivarié (Y_t, σ_t) . En effet, le filtre se base seulement

sur les observations Y_{t_k} pour les estimations. On remarque également que le filtre donne de relativement bonnes estimations des paramètres α , ξ et ρ . Nous remarquons également que plus la taille des simulations (n) augmente, et plus l'erreur quadratique moyenne diminue, ce qui est tout à fait normal. Pour pouvoir améliorer les estimations des paramètres β , κ et σ_2 , il faudrait prendre en compte les observations du processus de la volatilité σ_t dans le filtre ou encore changer de méthode de filtrage. Il est à remarquer que la difficulté à déterminer la valeur du paramètre κ revient au fait que plus ce paramètre est grand, moins la volatilité σ_t varie; elle devient presque constante. Par ailleurs, plus le paramètre ξ est petit, moins la volatilité σ_t varie; elle devient aussi dans ce cas presque constante. Dans chacun de ces cas, on retrouve le modèle de Black-Scholes avec volatilité constante!

Le graphique 7.1 montre les prévisions obtenues pour $n = 250$ si l'on connaît les valeurs des paramètres. On voit que la volatilité stochastique est très difficile à prévoir à court terme. Il est à noter que nous avons multiplié le bruit des observations par 100, sinon la différence entre les valeurs prévues et les valeurs réelles du S_t étaient trop petites.

Nous aimerais faire une remarque concernant les bons résultats des simulations obtenus dans l'article de Nielsen et al (2000). Il faut remarquer que dans le cas de l'information complète, ces auteurs supposent le processus de la volatilité connu et observable, ce qui n'est pas le cas en réalité. Deuxièmement, dans le cas de l'information incomplète, Nielsen (2000) suppose que $\kappa = 1$ et l'élimine des paramètres à estimer. Comme nous l'avons vu, c'est le paramètre κ qui est le plus difficile à estimer.

Donc, si Nielsen et all (2000) ont pu obtenir de si bons résultats, il faut les relativiser car ils se basent sur des suppositions qui ne correspondent pas à la réalité.

7.4 Application aux données financières

Dans cette section, nous appliquons le filtre gaussien du second ordre ainsi que la méthode du maximum de vraisemblance aux données financières que nous avons présentées au début du mémoire, à savoir celles des cotes de fermetures des titres présentés plus haut. La première valeur est choisie comme étant la valeur initiale du filtre. Les données sont constituées des cotes de fermetures de 4 titres entre le 9 Septembre 1991 et le 10 Septembre 2001, en utilisant un pas temporel égal à 1, indiquant que les valeurs des paramètres estimés sont des données quotidiennes. On initialise la matrice des covariances \mathbf{P} à la matrice identité \mathbf{I} . On considère que le processus des prix de nos 4 actifs peuvent être décrit à partir du modèle de Black-Scholes-Courtadon tel que nous l'avons présenté au paragraphe précédent. Nous avons fixé les valeurs de départ du filtre à partir des données de nos séries en nous basant sur le modèle de Black-Scholes. La valeur initiale du paramètre β a été fixé à $\hat{\beta} = s$, où s^2 est la variance des données $R_k = \log Y_{t_k} - \log Y_{t_{k-1}}$, $2 \leq k \leq N$. La valeur initiale du paramètre α a été choisie égale à $\hat{\alpha} = \bar{R} + \hat{\beta}^2/2$. Nous avons fixé les autres paramètres à $\kappa = 1$, $\xi = 1$ et $\rho = 0$. Les résultats des estimations sont présentés dans le tableau 7.2. Les graphiques 7.2, 7.3, 7.4 et 7.5 montrent les prévisions des prix et des

volatilités.

On voit que les prévisions des 4 titres sont excellentes. Les valeurs réelles se situent presque toutes dans la courbe de confiance, calculées en prenant $\hat{x}_{t_k|t_{k-1}}(1) \pm \sqrt{P_{t_k|t_{k-1}}(1, 1)}$ pour les bornes. Pour ce qui est de la volatilité, elle est estimée comme étant presque constante. Ceci est dû au fait que les valeurs estimées des paramètres κ des 4 titres sont élevées, ce qui conduit à un retour à la moyenne très rapide. Les bornes de prévisions de la volatilité sont données par $\hat{x}_{t_k|t_k}(2) \pm \sqrt{P_{t_k|t_k}(2, 2)}$. Comme les valeurs de κ sont grandes, il en résulte que $\hat{x}_{t_k|t_{k-1}}(2)$ est presque toujours constant, de même que $P_{t_k|t_k}(2, 2)$. Les équations suivantes, obtenues en résolvant en partie les systèmes d'équations différentielles (7.7) et (7.8)

$$\begin{aligned} \{\hat{x}_{t_k|t_{k-1}}\}_1 &= \{x_{t_{k-1}|t_{k-1}}\}_1 e^{\alpha\tau}, \\ \{\hat{x}_{t_k|t_{k-1}}\}_2 &= \beta + (\{x_{t_{k-1}|t_{k-1}}\}_2 - \beta)e^{-\kappa\tau}, \\ \{P_{t_k|t_{k-1}}\}_{2,2} &= \frac{\beta^2 \xi^2}{\kappa - \xi^2} \left(1 - e^{-(\kappa - \xi^2)\tau}\right) + \{P_{t_{k-1}|t_{k-1}}\}_{2,2} e^{-(\kappa - \xi^2)\tau} \\ &\quad + 2\beta(\{x_{t_{k-1}|t_{k-1}}\}_2 - \beta)e^{-(\kappa - \xi^2)\tau} \left(1 - e^{-\xi^2\tau}\right) \\ &\quad - \frac{\xi^2(\{x_{t_{k-1}|t_{k-1}}\}_2 - \beta)^2}{\kappa + \xi^2} \left(1 - e^{-(\kappa + \xi^2)\tau}\right), \end{aligned}$$

où $\tau = t_k - t_{k-1}$.

On voit alors que plus κ est grand, plus $\{\hat{x}_{t_k|t_{k-1}}\}_2$ tend vers β et plus $\{P_{t_k|t_{k-1}}\}_{2,2}$ tend vers $\frac{\beta^2 \xi^2}{\kappa - \xi^2}$. C'est pourquoi les valeurs de $\{P_{t_k|t_{k-1}}\}_{2,2}$ sont presque constantes.

Nous remarquons dans le dernier graphique de la figure (7.4) que les premières valeurs sont trop grandes par rapport aux suivantes.

Titre	α	β	κ	ξ	σ_e	ρ
GM	0.001352	0.017745	4.310262	1.051699	0.105257	0.119850
Coca-Cola	0.000039	0.022581	47.089440	1.595793	0.068206	-0.999773
Microsoft	0.001437	0.005639	1.019636	1.381118	0.801691	0.026826
Mobil	0.004758	0.002426	6.564155	3.557712	0.297545	-0.271185

Tableau 7.2: Estimation des paramètres des séries des cotes de fermetures des titres de GM, Coca-Cola, Microsoft et Mobil.

7.5 Conclusion

Dans le présent mémoire, nous avons passé en revue les différentes méthodes utilisées dans l'estimation de la volatilité. Nous avons d'abord commencé par une comparaison des méthodes classiques, à savoir l'estimation de la volatilité historique et implicite dans le cadre d'une modélisation à volatilité constante. Nous avons montré les avantages et les inconvénients de chacune de ces deux méthodes avec des illustrations numériques, avant de démontrer qu'aucune des deux méthodes ne correspond à la réalité du marché; à savoir que la volatilité suit son propre processus de diffusion caché. Par la suite, nous avons introduit une méthode de filtrage non linéaire qui permet d'estimer les paramètres du processus de la volatilité dans le cadre d'une modélisation à volatilité stochastique. Durant l'implémentation du filtre, nous avons fait face à plusieurs problèmes numériques et computationnels. Il nous fallait d'abord spécifier un ensemble de paramètres à estimer qui ne fasse pas en sorte que l'estimation de la matrice des covariances ne tend vers l'infini.

Pour se prémunir de ce genre de problème, il nous faut vérifier l'état de la matrice des covariances à chaque étape de l'estimation, et introduire une fonction de pénalité à la fonction à la fonction de vraisemblance tel que suggéré dans l'article de (Madsen et Melgaard, 1991). Chose que nous n'avons pas entrepris de faire pour ne pas prolonger les temps de calcul qui sont déjà long. Les calculs ont été réalisé sur un ordinateur Pentium III, cadencé à 900 MHz avec une mémoire cache de 256 Mo. Nous avons choisi d'implémenter le filtre avec le langage Matlab (v6.12) pour la facilité d'utilisation de l'environnement de développement ainsi que pour son aspect graphique. Le problème avec le langage Matlab est la vitesse d'exécution des procédures et spécialement l'exécution des boucles *FOR* et *WHILE*. Un autre problème à soulever avec le langage Matlab est que nous faisons confiance aux fonctions disponibles dans le langage sans savoir comment les fonctions de minimisation que nous avons utiliser *FMINSEARCH* ont été programmé. Une autre procédure de minimisation devrait être utilisée, car nous doutons que la procédure de minimisation offerte avec le langage Matlab donne une bonne estimation de la matrice hessienne. Une meilleure estimation de la matrice hessienne nous permettrait d'obtenir une meilleure estimation de la variance et de la matrice de corrélation. D'autres améliorations peuvent être faites quand nous avons à estimer les paramètres sur des données réelles dans le but d'obtenir des estimations plus réalistes. En appliquant le filtre à des données réelles, le pas de discréétisation temporel doit être égal à $\frac{1}{365}$. Et comme nous savons qu'il y a exactement 250 jours ouvrables dans une année, on peut déterminer avec précision les 115 où il n'y a pas de transaction boursières, et faire en sorte que le filtre ne fasse de mise à jour des

équations du filtres que s'il reconnaît un jour comme étant un jour ouvrable. Nous pouvons aussi appliquer le filtre à d'autres champs tel que l'évaluation des options sous volatilité stochastique. Ceci n'est qu'un bref aperçu des suggestions pour les travaux futurs. Une autre amélioration que nous pourrions étudier dans le futur est la possibilité de considérer un autre processus stochastique pour la dérive (“*drift*”). Cette possibilité d'estimer des entités inobservables sur le marché financier ouvre la voie à l'invention de nouveaux outils financiers.

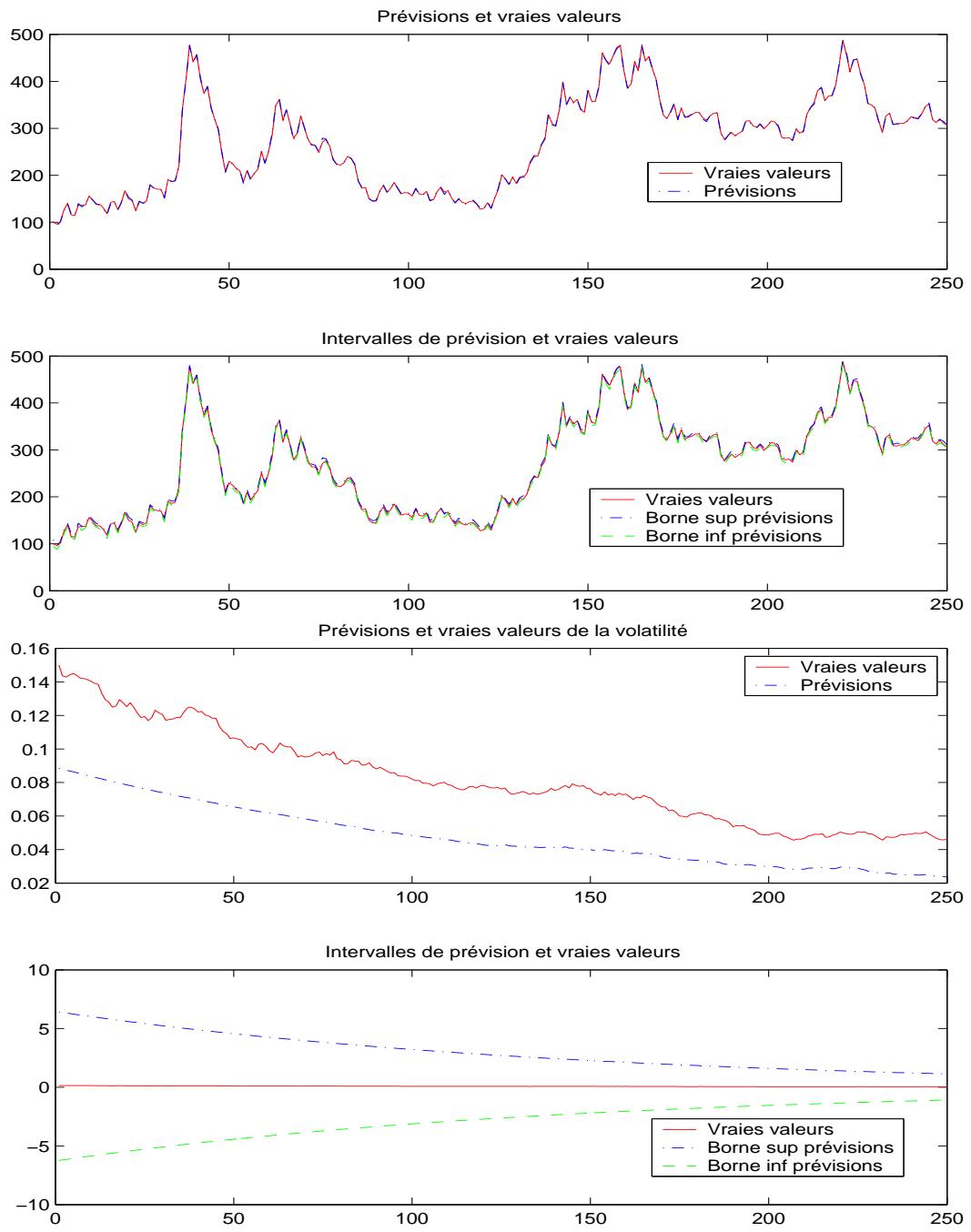


Figure 7.1: Prévisions avec paramètres $\alpha = 0.00198$, $\beta = 0.01$, $\kappa = 0.0071$, $\xi = 0.016$, $\sigma_e = 1.95$, et $\rho = 0.5$.

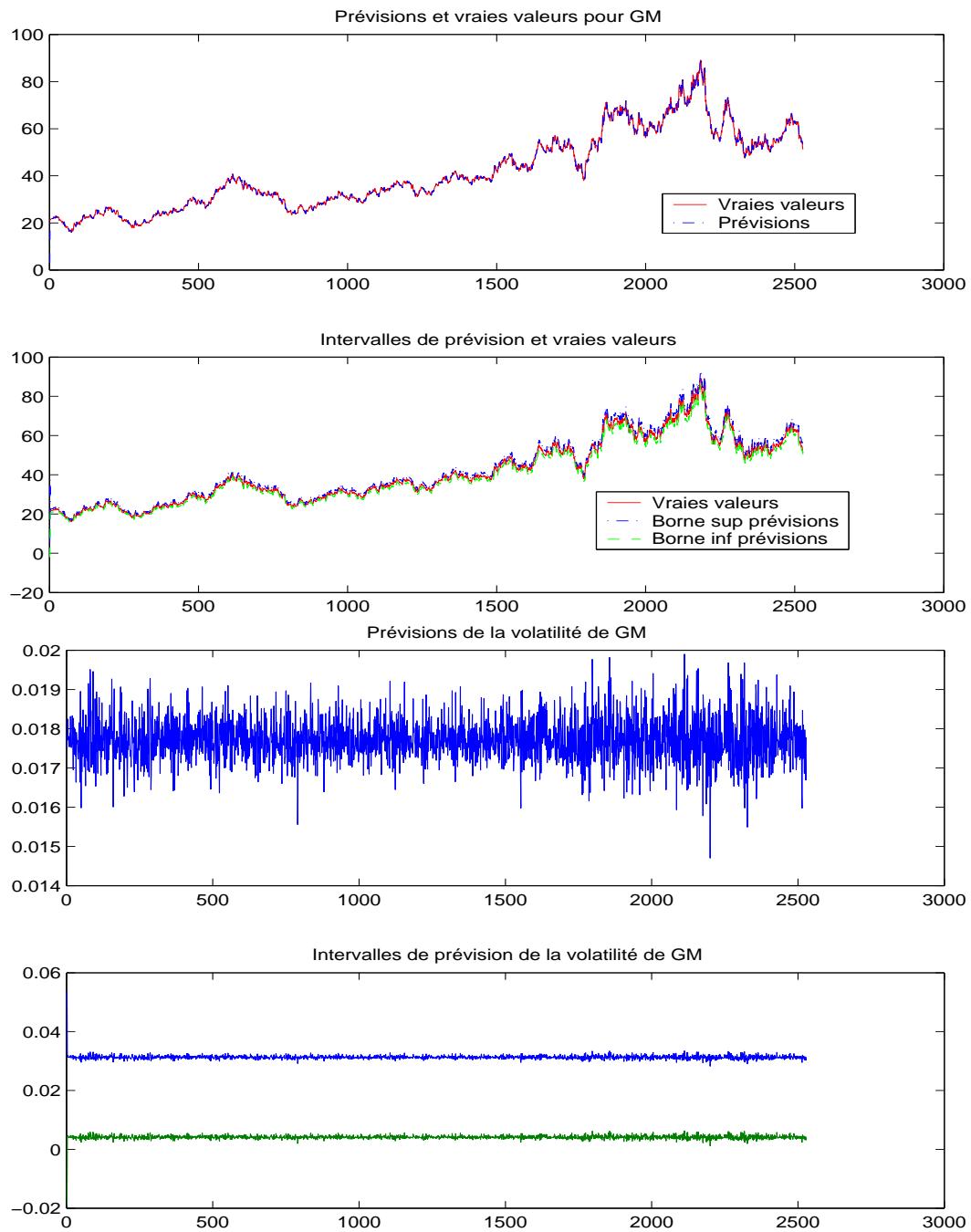


Figure 7.2: Prévisions du prix de GM et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)

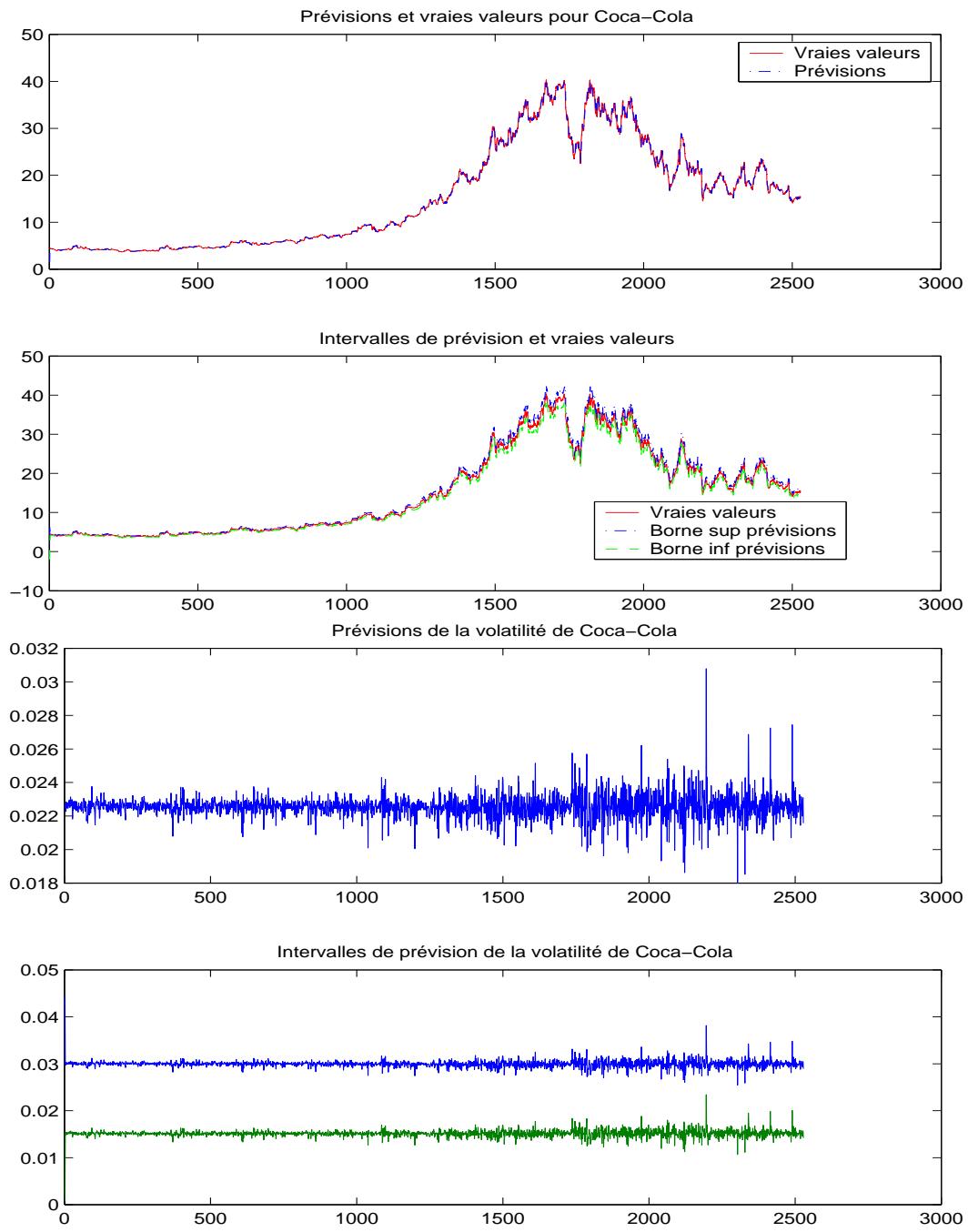


Figure 7.3: Prévisions du prix de Coca-Cola et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)

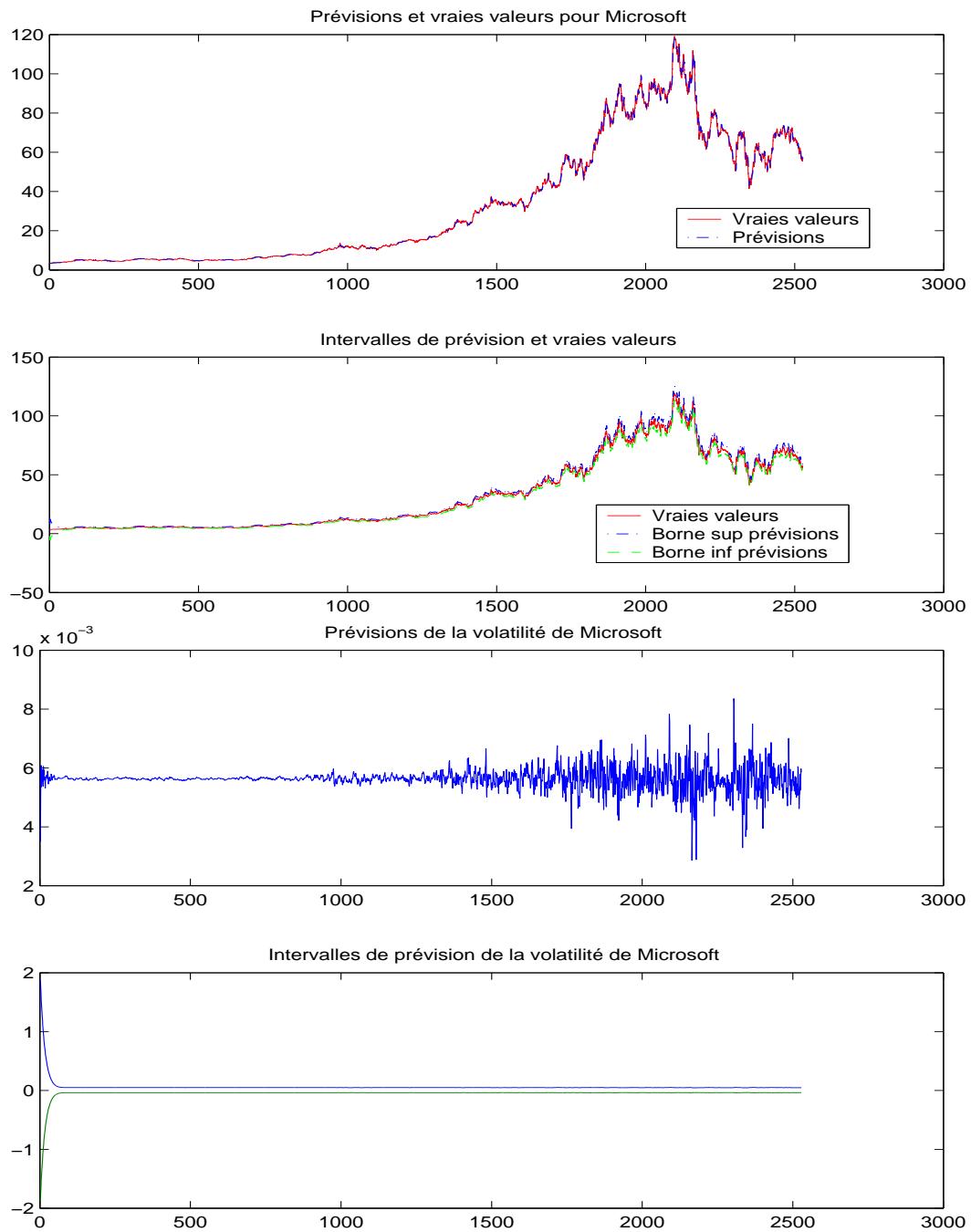


Figure 7.4: Prévisions du prix de Microsoft et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)

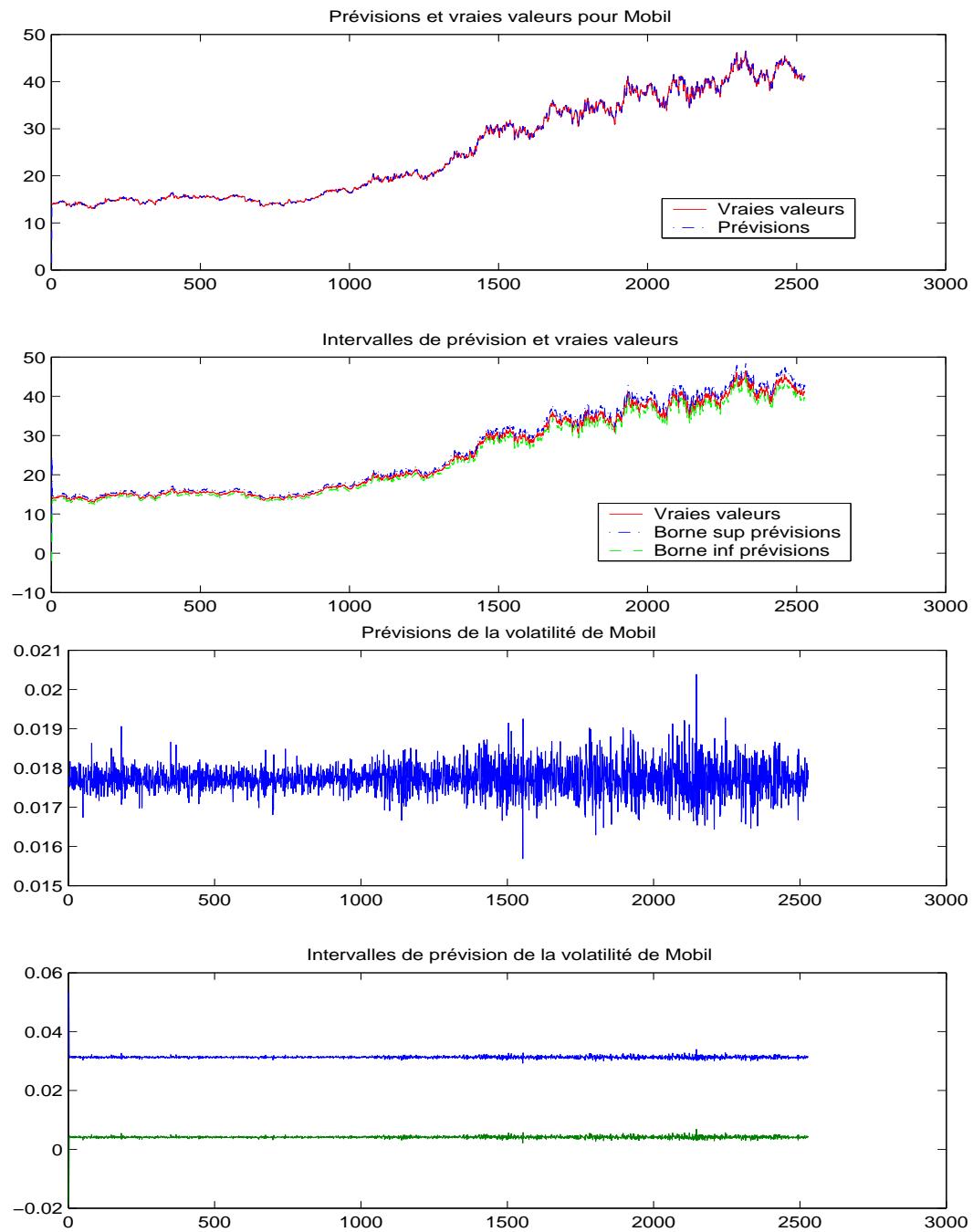


Figure 7.5: Prévisions du prix de Mobil et de sa volatilité d'après le modèle de Black-Scholes-Courtadon.)

Appendice A

Mouvement brownien

A.1 Définition du mouvement brownien

Nous attribuons au mathématicien Norbert Wiener l'analyse rigoureuse des mathématiques concernant le mouvement brownien (1863) et c'est pourquoi ce processus est aussi connu sous le nom de processus de Wiener.

Définition: Un *mouvement brownien standard* $W_t : t \geq 0$ est un processus stochastique adapté construit sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$ tel que:

- (i) $\forall \omega \in \Omega, W_0(\omega) = 0,$
- (ii) $\forall 0 \leq t_0 \leq t_1 \leq \dots \leq t_k,$ les variables aléatoires $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_k} - W_{t_{k-1}}$ sont indépendantes,

- (iii) $\forall s, t \geq 0$ tels que $s < t$, la variable aléatoire $W_t - W_s$ est de loi normale de moyenne 0 et de variance $t - s$,
- (iv) $\forall \omega \in \Omega$, la trajectoire $t \mapsto W_t(\omega)$ est continue.

En général, la filtration utilisée est $\mathbb{F} = \mathcal{F}_t : t \geq 0$ où $\mathcal{F}_t = \sigma\{W_s : 0 \leq s \leq t\}$ est la plus petite tribu pour laquelle les variables aléatoires W_s , $0 \leq s \leq t$, sont mesurables.

A.2 Propriétés du mouvement brownien

Lemme A.2.1 *Soit $W_t : t \geq 0$, un mouvement brownien standard. Alors, les processus suivants sont aussi des mouvements browniens standards:*

(i) $W_{t+s} - W_s : t \geq 0$, s fixé (homogénéité dans le temps)

(ii) $-W_t : t \geq 0$ (symétrie)

(iii) $cW_{\frac{t}{c^2}} : t \geq 0$ (rééchelonnement du temps)

(iv)

$$W_t^* = \begin{cases} 0 & \text{si } t = 0 \\ tW_{\frac{1}{t}} & \text{si } t > 0 \end{cases}$$

(inversion du temps)

Lemme A.2.2 *Le mouvement brownien est une martingale*

Lemme A.2.3 *Le mouvement brownien est un processus markovien*

Lemme A.2.4 $\forall \omega \in \Omega$, la trajectoire $t \rightarrow W_t(\omega)$ est nulle part différentiable.

A.3 Intégrale stochastique et calcul d'Itô

A.3.1 Définition de l'intégrale stochastique par rapport à une martingale à temps discret

Soit $\mathbf{M} = \mathbf{M}_n : n \in \mathbb{N}$, une martingale à temps discret construite sur l'espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$, $\mathbb{F} = \mathcal{F}_n : n \in \mathbb{N}$. Pour tout processus \mathbb{F} -prévisible $X = X_n : n \in \mathbb{N}$ nous définissons l'intégrale stochastique de X par rapport à M par:

$$\forall n \in \mathbb{N} \text{ et } \forall \omega \in \Omega,$$

$$\left(\int_0^n X_m dM_m \right) (\omega) = \sum_{m=1}^n X_m(\omega)(M_m(\omega) - M_{m-1}(\omega)).$$

Interprétation: Supposons que M est une martingale représentant l'évolution du prix actualisé d'une part d'un titre boursier. Supposons aussi que le nombre de part détenues à chaque instant est modélisé par le processus stochastique X . Alors, l'intégrale stochastique $\int_0^n X_m dM_m$ représente la valeur actualisée de la fortune de l'investisseur au temps n . Soulignons que, contrairement à l'intégrale de Stieltjes qui est déterministe, $\int_0^n X_m dM_m$ est une variable aléatoire.

A.3.2 Définition de l'intégrale stochastique par rapport au mouvement brownien

Soit $W_t : t \geq 0$ un mouvement brownien standard construit sur un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$, $\mathbb{F} = \mathcal{F}_n : n \in \mathbb{N}$.

Soit $X_t : t \geq 0$ un processus stochastique prévisible. On peut penser à définir l'intégrale stochastique de X par rapport à W , trajectoire par trajectoire, en utilisant une généralisation de l'intégrale de Stieltjes:

$$\forall \omega \in \Omega, \int X(\omega) dW(\omega).$$

Cela serait possible si les trajectoires du mouvement brownien W étaient à variation bornée, c'est à dire qu'elles pourraient s'exprimer comme une différence de deux fonctions non décroissantes. Or, comme les trajectoires du mouvement brownien ne sont pas à variation bornée, il n'est pas possible d'utiliser cette approche. Nous devons définir l'intégrale stochastique par rapport au mouvement brownien globalement, c'est à dire comme un processus stochastique en lui-même, et non pas trajectoire par trajectoire.

Les processus stochastiques de base

Nous appelons X un processus stochastique de base si X admet la représentation

$$X_t(\omega) = C(\omega) \delta_{(a,b]}(t),$$

où $a < b \in \mathbb{R}$ et C est une variable aléatoire \mathcal{F}_a -mesurable de carré intégrable, c'est à dire que $E_P[C^2] < \infty$. Remarquons que ce processus est adapté à la

filtration \mathbb{F} . En fait, X est plus que \mathbb{F} -mesurable, elle est \mathbb{F} -prévisible, mais comme la notion de processus prévisible est plus délicate à définir lorsque nous travaillons avec des processus à temps continu. Notons tout de même que les processus adaptés à trajectoires continues sont prévisibles. Intuitivement, si X_t représente le nombre de parts d'un titre détenues au temps t . alors $X_t(\omega) = C(\omega)\delta_{(a,b]}(t)$ signifie qu'immédiatement après l'annonce des prix au temps a et sur la base de l'information disponible au temps a , nous achetons $C(\omega)$ parts du titre que nous conservons jusqu'au temps b .

L'intégrale stochastique de X par rapport au mouvement brownien est définie par

$$\begin{aligned} \left(\int_0^t X_s dW_s \right) (\omega) &= C(\omega)(W_{t \wedge b}(\omega) - W_{t \wedge a}(\omega)) \\ &= \begin{cases} 0 & \text{si } 0 \leq t \leq a \\ C(\omega)(W_t(\omega) - W_a(\omega)) & \text{si } 0 \leq t \leq b \\ C(\omega)(W_b(\omega) - W_a(\omega)) & \text{si } b < t \end{cases} \end{aligned}$$

Les processus stochastiques simples

Nous appelons X un processus stochastique simple si X est une somme finie de processus de base:

$$X_t(\omega) = \sum_{i=1}^n C_i(\omega)\delta_{(a_i, b_i]}(t),$$

Ce processus est \mathbb{F} -prévisible comme sommation de processus prévisibles. L'intégrale stochastique par rapport au mouvement brownien est définie

comme la somme des intégrales stochastiques des processus de base constituant X :

$$\left(\int_0^t X_s dW_s \right) (\omega) = \sum_{i=1}^n \int_0^t C_i(\omega) \delta_{(a_i, b_i]}(s) dW_s$$

Il est possible d'étendre la classe de processus pour lesquels l'intégrale stochastique peut être définie. Cela fait intervenir les martingales locales ainsi que les semi-martingales[Durrett, 1996].

A.3.3 Calcul d'Itô

Le calcul d'Itô permet d'introduire un calcul différentiel sur les intégrales stochastiques, et l'outil essentiel en est la formule d'Itô. La formule d'Itô donne, en particulier, la façon de différencier $t \mapsto f(W_t)$ si f est une fonction deux fois continûment dérivable. Commençons par préciser la définition de la classe de processus pour laquelle on peut énoncer la formule d'Itô.

Définition A.3.1 *Soient $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ un espace probabilisé muni d'une filtration et $(W_t)_{t \geq 0}$ un \mathcal{F}_t -mouvement brownien. On appelle processus d'Itô, un processus $(X_t)_{0 \leq t \leq T}$ à valeurs dans \mathbb{R} tel que:*

$$\mathbf{P} p.s \forall t \leq T \quad X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s,$$

Théorème A.3.1 Formule d'Itô: *Soit $(X_t)_{0 \leq t \leq T}$ un processus d'Itô:*

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s,$$

et f une fonction deux fois continûment différentiable, on a:

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X, X \rangle_s$$

où, par définition:

$$\langle X, X \rangle_t = \int_0^t H_s^2 ds,$$

et:

$$\int_0^t f'(X_s) dX_s = \int_0^t f'(X_s) K_s ds + \int_0^t f'(X_s) H_s dW_s.$$

Pour une démonstration complète, voir [Karatzas and Shreve, 1996]

A.3.4 Équations différentielles stochastiques

Si X est un processus d'Itô, alors le lemme d'Itô nous permet de trouver l'équation différentielle stochastique satisfaite par $f(X)$. Or, nous voulons l'inverse, c'est à dire qu'étant donné une équation différentielle stochastique, nous voudrions trouver un processus stochastique la satisfaisant. Tout comme dans le cas stochastique, il existe des équations qui n'ont pas de solution et d'autres qui possèdent une infinité de solution. Le résultat suivant présente des conditions suffisantes à l'existence d'une solution unique:

Théorème A.3.2 Soit l'équation différentielle stochastique

$$dX_t = b(X(t), t) dt + a(X(t), t) dW(t) \quad (7)$$

Si $b : \mathbb{R} \times [0, \infty) \rightarrow \mathbb{R}$ et $a : \mathbb{R} \times [0, \infty) \rightarrow \mathbb{R}$ sont des fonctions continues telles qu'il existe une constante $K < \infty$ pour laquelle

$$(i) \quad |b(x, t) - b(y, t)| + |a(x, t) - a(y, t)| \leq K|x - y|, \quad \forall t \geq 0$$

$$(ii) \ |b(x, t)|^2 + |a(x, t)|^2 \leq K^2(1 + x^2), \ \forall t \geq 0$$

$$(iii) \ E[X_0^2] < \infty$$

alors il existe une solution unique à l'équation (7)

Il n'existe pas d'algorithme permettant la résolution de toutes les équations différentielles stochastiques. Il existe des équations différentielles stochastiques dont nous savons qu'elles possèdent une solution unique, mais que nous ne sommes pas capable d'exprimer analytiquement. Cependant, nous pouvons utiliser des approximations ou des techniques numériques pour nous aider à mieux comprendre le comportement de processus stochastiques caractérisés par des équations différentielles stochastiques.

Appendice B

Algorithme de filtrage et d'estimation

L'algorithme de filtrage et d'estimation utilisé au dernier chapitre se présente de la manière suivante:

1. N=nombre de données à simuler
2. Lancement du programme Simulation
 - (a) On fixe les paramètres $\alpha, \beta, \kappa, \theta, \sigma_e$ et ρ
 - (b) On génère le processus des logarithmes la volatilité σ_t
 - (c) On génère le processus des logarithmes des prix S_t
 - (d) On détermine le processus observé en temps discret à partir de S_t :

$$Y_{t_k} = S_{t_k} + e_{t_k}, \quad e_{t_k} \sim N(0, \sigma_e^2)$$

3. Initialisation du filtre gaussien du second ordre
 - (a) Initialisation de la moyenne conditionnelle
 - (b) Initialisation de la matrice des covariances conditionnelles
4. Calcul des équations de propagation et de mise à jour du filtre
5. Calcul de la matrice de gain du filtre G
6. Calcul de la fonction de maximum de vraisemblance
7. Minimisation de la fonction de maximum de vraisemblance
 - (a) On obtient une estimation des paramètres $\alpha, \beta, \kappa, \theta, \sigma_e$ et ρ
8. Processus filtrés X^{fil} et σ_t^{fil}
9. Graphique des processus simulés et ceux obtenus à partir du filtre

Appendice C

Programmes Matlab

```
% simulation2.m permet la simulation du processus du log des prix
% ainsi que celui du log de la volatilit\'{e} stochastique.
% Nous utilisons la m\'{e}thode d'Euler pour la discr\'{e}tisation des EDS correspondantes.
% Nous consid\'{e}rons le mod\`{e}le de Black&Scholes-Courtadon.
% SAIMI NABIL
% Universit\'{e} du Qu\'{e}bec a Trois-Rivieres- HEC
% Montr\'{e}al 2001-2002
% Les \'{e}quations simul\'{e}es sont:
% 1) dS_t=\alpha S_t dt + \sigma_t S_t dW1
% 2) d\sigma_t= \kappa(\beta - \sigma_t)dt + \theta \sigma_t dW2
% que nous pouvons transformer avec le changement de variable suivant:
% X_t= log S_t et V_t = log \sigma_t
% ce qui donne les \'{e}quations suivantes:
% 1) dX_t = (\alpha-1/2\exp(2*V_t))dt + \exp(V_t) dW1
% 2) dV_t = \kappa(\beta - \exp(V_t))\exp(-2*V_t)-(1/2 \theta^2)dt + \theta dW2
% Initialisation des constantes
% Iter = 10;
% n = 1000;
function Y = simulation2(n,Iter);
```

```

k = 100; %sampling frequency
T = n/250; % ann\'ees
h = 1/(250*k); % pas temporel
N = n*k; % nombre de pas
% t = (0:h:n); % t est le vecteur [0 1h 2h 3h ... Nh]

alpha = 0.00198;
beta = 0.01;
kappa = 0.0071;
xi = 0.016;
sigma_e = 0.0195;
rho = 0.5;

X=zeros(Iter,N+1);
X(:,1)=log(100.0); % Valeur initiale du processus du log des prix
V=zeros(Iter,N+1);
V(:,1)=log(0.15); % Valeur initiale du processus du log de la volatilit\'{e}

%%%%%%%%%%%%%
% le vecteur V repr\'{e}sente le log du processus de la volatilit\'{e}
% le vecteur X repr\'{e}sente le log du processus des prix
% pour s'assurer que les prix simul\'{e}s soient positifs.
%
%
%
e1=randn(Iter,N);
e2=randn(Iter,N);
e3=randn(Iter,n)*sigma_e;
for j=1:Iter
for i=1:N % Discr\'{e}tisation du processus du log des prix Eqn #49
    V(j,i+1)= V(j,i) + [kappa*(beta*exp(-V(j,i))-1)-xi^2/2]*h
        + [rho*xi*sqrt(h)*e1(j,i)] + [sqrt(1-rho^2)*xi*e2(j,i)*sqrt(h)] ;
    X(j,i+1)= X(j,i) + [alpha-exp(2*V(j,i))/2]*h + [exp(V(j,i))*e1(j,i)*sqrt(h)];

```

```

    end;
end;

Y=zeros(Iter,n);

for i=1:n
    Y(:,i)= exp(X(:,k*(i-1)+1))+e3(:,i);
end;

% ----- Fin du programme-----

%%%%% GSOF4.m contient l'impl\'ementation du filtre gaussien du second ordre
% ainsi que la proc\'edure du maximum de vraisemblance pour l'estimation
% des parametres du systeme d'\equations.
% NABIL SAIMI, Janvier 2002
% Universit\'e du Qu\'ebec a Trois-Rivieres.
% M.Sc Applied Mathematics and Computer Science
%%%%%
function f = GSOF4(par,Y,n,k);

%h = 1/(250*k);
h = 1/k;
%parametres du modele de Black-Scholes-Courtadon
alpha=par(1);
beta= exp(par(2));
kappa= exp(par(3));
xi=exp(par(4));
sigma_e=exp(par(5));
rho= tanh(par(6));

*****
% Initialisation des parametres:
% alpha=0.00198;

```

```

% %kappa=0.0071;
% beta=0.01;
% xi=0.016;
% sigma_e=0.0195;
% rho=0.5;
%*****


% Initialisation de la moyenne conditionnelle et la covariance
% de l'\'{e}tat du systeme: pre pour A priori et post pour
% A posteriori.
preP=zeros(2,2);      % Covariance a priori
postP=zeros(2,2);      % Covariance a posteriori
preMoy=zeros(2,n);      % Moyenne conditionnelle a priori
postMoy=zeros(2,n);      % Moyenne conditionnelle a posteriori
resid=zeros(1,n);      %
R=zeros(1,n);
%*****


%simulation2 ;      %fonction de simulation du processus des prix
%et de la volatilit\'{e}.
%fprintf('simulation termin\'{e}e!');




% Valeurs initiales du filtre
postP = [1 0; 0 1];      % Matrice des covariances.

postMoy(:,1) = [Y(1); beta]; % Moyennes conditionnelles.

ptil=zeros(2,2);

DUM = zeros(2,2);
for i=2:n
    % Equations d'\'{e}volution page15 Nielsen
    %Eqs 52 et 53

    s = postMoy(:,i-1);
    D = postP;

    for j=1:k

```

```

preMoy(:,i)= s + h*ff(s,alpha,beta,kappa);
preP = D + h*G(D,alpha,kappa) + h*Ptilde(s,D,xi,rho);
s = preMoy(:,i);
D = preP;
end;

% Nielsen, Eq 18
R(i)= preP(1,1) + sigma_e^2;
%Equations de mise 'a jour (Eq 55, Nielsen)
postMoy(:,i)=preMoy(:,i)+ preP(:,1)*(Y(i)-preMoy(1,i))/R(i);
%%%%%%%%%%%%%
%Eq 56, Nielsen
postP = inv(inv(preP)+ [1/sigma_e^2 0; 0 0]);
resid(i)= Y(i)-preMoy(1,i);

end;

MV=0;
for i= 2:n
    MV=MV + 1/2*[log(R(i))+ resid(i)^2/R(i)];
end;

f=MV;

% ----- Fin du programme-----
%%%%%
% minimisa4.m permet la minimisation de la fonction du maximum de
% vraisemblance a partir des r\'esultats du filtre gaussien du
% second ordre.

```

```

% NABIL SAIMI, Janvier 2002
% Universit\{e} du Qu\{e}bec a Trois-Rivieres.
% Master in Applied Mathematics and Computer Science
%%%%%%%%%%%%%
% programme pour l'estimation des variance des bruits R et Q
% L'id\{e}e est de minimiser L(y,R,Q);
%
%
tic
warning off
n = 500;
Iter = 2;
param = zeros(1,6);
par0 = [0.00198 log(0.01) log(0.0071) log(0.016) log(0.0195) atanh(0.5)]';
fid = fopen('s500.2.txt','w');

k = 50;

Y = simulation2(n,Iter);

for i = 1:Iter
    Iteration = i

%par0 = 0.01+zeros(1,6)';
%par0 = [0.00198 log(0.01) log(0.0071) log(0.016) log(0.0195) atanh(0.5)]';

%
% A=(-1)*eye(6);
% b = zeros(1,6)';
% lb=-1*ones(6,1);
% ub=[1 1 3 1 1 1]';

%
%options = optimset('GradObj','on');%'Hessian','on');
%options=optimset('MaxFunEvals',1000);%, 'Hessian','on');

```

```

%options=optimset('largescale','on','MaxFunEvals',1000,'MaxIter',100,'Display','iter');%
%g=fmincon('GSOF2' ,par0,A,b,[],[],lb,ub,[],options,Y);
options=optimset('largescale','on','MaxFunEvals',1000,'MaxIter',100);%
g =fminsearch('GSOF4' ,par0,options,Y(i,:),n,k);
%options=optimset('Display','iter');%
%[g,fval,exitflag,output] =fminsearch('GSOF4' ,par0, options, Y(i,:),n,k)
param(1) = g(1);
param(2) = exp(g(2));
param(3) = exp(g(3));
param(4) = exp(g(4));
param(5) = exp(g(5));
param(6) = tanh(g(6));
fprintf(fid,' %f %f %f %f %f %f \n',param);

end;
fclose(fid);

toc
% ----- Fin du programme-----

```

```

%%%%%
% minimisa.m permet la minimisation de la fonction du maximum de %
% vraisemblance a partir des r\'{e}sultats du filtre gaussien du      %
% second ordre.          %
% NABIL SAIMI, Janvier 2002          %
% Universit\'{e} du Qu\'{e}bec a Trois-Rivieres.          %
% Master in Applied Mathematics and Computer Science          %
%%%%%
% programme pour l'estimation des variance des bruits  R et Q
% L'id\'{e}e est de minimiser L(y,R,Q);
%
%
%function g=minimisa5;
%tic
%fid = fopen('coke.par','w');

```

```

load coke.txt;
Y = coke;

n = length(Y);

k=50;

dum = zeros(n-1,1);

for i = 1:(n-1)
    dum(i) = log(Y(i+1)/Y(i));
end;

beta = sqrt(var(dum));
a = mean(dum);
alpha = a + 0.5*beta*beta;

param = zeros(1,6);
par = [ alpha log(beta) 0 0 0 0]';

%par0 = 0.01+zeros(1,6)';
%par0 = [0.00198 log(0.01) log(0.0071) log(0.016) log(0.0195) atanh(0.5)]';

% A=(-1)*eye(6);
% b = zeros(1,6)';
% lb=-1*ones(6,1);
% ub=[1 1 3 1 1 1]';

%options = optimset('GradObj','on');%'Hessian','on');
%options=optimset('MaxFunEvals',1000);%, 'Hessian','on');
options=optimset('largeScale','on','MaxFunEvals',1000,'MaxIter',200,'Display','iter');
%g=fmincon('GSOF2' ,par0,A,b,[],[],lb,ub,[],options,Y);
g=fminsearch('GSOF4' ,par,options,Y,n,k);

```

```

param(1) = g(1);
param(2) = exp(g(2));
param(3) = exp(g(3));
param(4) = exp(g(4));
param(5) = exp(g(5));
param(6) = tanh(g(6));
fprintf(fid, ' %f %f %f %f %f %f \n',param);

fclose(fid);

%%%%%%%%%%%%%%%
%h = 1/(250*k);
h = 1/k;
% Initialisation des parametres:

alpha=param(1);
kappa=param(3);
beta=param(2);
xi=param(4);
sigma_e=param(5);
rho=param(6);
*****


% Initialisation de la moyenne conditionnelle et la covariance
% de l'\{e}tat du systeme: pre pour A priori et post pour
% A posteriori.
preP=zeros(2,2);      % Covariance a priori
postP=zeros(2,2);      % Covariance a posteriori
preMoy=zeros(2,n);    % Moyenne conditionnelle a priori
postMoy=zeros(2,n);    % Moyenne conditionnelle a posteriori
postPri=zeros(1,n);    % Valeur
prediPri=zeros(1,n);

```

```

prediMoy=zeros(1,n);
resid=zeros(1,n);      %
R=zeros(1,n);

Var=zeros(2,n);
%*****%
% Valeurs initiales du filtre
postP = [1 0; 0 1];      % Matrice des covariances.

Var(:,1)= [postP(1,1); postP(2,2)];

postMoy(:,1) = [Y(1) ; 0]; % Moyennes conditionnelles.

ptil=zeros(2,2);

DUM = zeros(2,2);

for i=2:n
    % Equations d'évolution page15 Nielsen
    %Eqs 52 et 53

    s = postMoy(:,i-1);
    D = postP;

    for j=1:k
        preMoy(:,i)= s + h*ff(s,alpha,beta,kappa);
        preP = D + h*G(D,alpha,kappa) + h*Ptilde(s,D,xi,rho);
        s = preMoy(:,i);
        D = preP;
    end;

    % Nielsen, Eq 18

```

```

R(i)= preP(1,1) + sigma_e^2;
%Equations de mise 'a jour (Eq 55, Nielsen)
postMoy(:,i)=preMoy(:,i)+ preP(:,1)*(Y(i)-preMoy(1,i))/R(i);
%%%%%%%%%%%%%%%
%Eq 56, Nielsen
postP = inv(inv(preP)+ [1/sigma_e^2 0; 0 0]);

Var(:,i)= [preP(1,1); postP(2,2)];

end;

t = [1:n]';

subplot(2,1,1);
plot(t ,Y, '-r', t,preMoy(1,t), '-.b');
h = legend('Vraies valeurs','Pr\'{e}visions',1);
title('Pr\'{e}visions et vraies valeurs pour Coca-Cola');
subplot(2,1,2);
plot(t,Y, '-r', t, preMoy(1,t)+2*sqrt(Var(1,t)), '-.b',t, preMoy(1,t)-2*sqrt(Var(1,t)), '--g');
h = legend('Vraies valeurs','Borne sup pr\'{e}visions','Borne inf pr\'{e}visions',1);
title('Intervalles de pr\'{e}vision et vraies valeurs');
figure;
%
t = [2:n]';

subplot(2,1,1);
plot(t, postMoy(2,t));
title('Pr\'{e}visions de la volatilit\'{e} de Coca-Cola');
subplot(2,1,2);
plot(t, postMoy(2,t)+2*sqrt(Var(2,t)),t, postMoy(2,t)-2*sqrt(Var(2,t)));
title('Intervalles de pr\'{e}vision de la volatilit\'{e} de Coca-Cola');

%
toc
% ----- Fin du programme-----

```

```

% wiener_forward.m initialise l'\'{e}quation avant de Kolmogorov
% pour un processus de Wiener

Tf = 10;
Tfixed=10;
nx = 150;           % nombre de points dans l'espace
nt = 100;           % nombre de pas temporels
dx = 1/nx;
dt = Tf/nt;
x = (0:dx:1)';
t = (0:dt:Tf);
xc = (dx/2:dx:1-dx/2)';

% Initialisation des limites spatiales
xmin = -10;
xmax = 10;

p = xmax-xmin;      % longueur de l'intervalle
xp = xmin + p*x;    % nombre de points dans l'intervalle des x
xcp = xmin + p*xc;  % points centr\'{e}s dans l'intervalle des x

% Les coefficients a, b et c sont initialis\'{e}s:
sigma = 1;           % coefficient de diffusion
a = (sigma^2)*ones(size(xcp))/2;  % coefficient constant de la diffusion
b = zeros(size(xcp));           % d\'{e}rive zero (drift)
c = b;

% La condition initiale est mise ici pour
% etre concentr\'{e} au centre de l'intervalle

u0 = zeros(size(xp));          % densit\'{e} de X(0)
[junk,imid]=min(abs(x-1/2));  % trouve les points milieux
u0(imid) = 1;
% sum(u0)=1 pour normalisation

kolm_avant           % r\'{e}soud l'\'{e}quation avant de Kolmogorov

```

```

% -----
% Ce programme donne une solution de l'\'{e}quation avant de Kolmogorov,
% qui permet de retrouver la densit\'{e}
% de la solution d'une \'{e}quation diff\'{e}rentielle stochastique(SDE)
% Nabil Saimi, Janvier 2002.
% Maitrise en Math\'{e}matiques et Informatique Appliqu\'{e}s
% Sujet: Estimation de la volatilit\'{e} et filtrage non-lin\'{e}aire.
% Les valeurs initiales doivent \^etre initialis\'{e}es par l'utilisateur.
% -----



a=a/(p^2);    % Mise en \'{e}chelle des co\''{e}fficients pour l'algorithme.
b=b/p;
u0=u0*nx/p;  % La normalisation donne: sum(u0)*p/nx=1

% Initialisation des matrices de masse et de rigidit\'{e}, ainsi
% que de la matrice finale "A" tel que u'(t) = Au(t).
A1 = make_A1(a);
A2 = make_A2(b);
A3 = make_A3(c);
M = make_A3(ones(size(c)));
A = M\ (A1 + A2 + A3);
emax = max(real(eig(A)));
A = A-emax*eye(size(A));

% Mise a jour temporelle(En arriere, par un algo de RK de 2eme ordre)

G = inv(eye(size(A)) - dt*A + (dt*A)^2/2 );
u = u0;
U = [u0];
for j=1:nt-1;
    u = G*u;
    U = [U,u];
end;
U = U(:,1:nt);
t = t(:,1:nt);

```

```

% Affichage de l'\\'{e}volution de la solution dans le temps

clf

subplot(2,1,1)
mesh(t,xp,U);
axis([0 Tf xmin xmax min(min(U)) max(max(U(:,5:nt)))]);
title('Evolution de la densit\'{e} dans le temps');
drawnow;

subplot(2,2,3)
ufixed = expm(A*Tfixed)*u0;
plot(xp,ufixed);
title(['Densit\'{e} au temps ' int2str(Tf)]);
grid on;

subplot(2,2,4)
secmom = (xp .* xp)' * U;
plot(t,secmom)
title('Deuxieme moment de X_{t}')

% ----- Fin du programme-----

```

```

%%%%%
% Cette fonction permet de calculer la valeur d'une option
% suivant le mod\'{e}le de Black&Scholes
%
% Nabil Saimi Copyright 2001
% Universit\'{e} du Qu\'{e}bec a Trois-Rivi\{e}res
%%%%%

function y=BS(S,X,vol,time,r,indice)

d1=(log(S/X)+(r+vol*vol/2)*time)/(vol*time^0.5);
d2=d1-vol*time^0.5;

```

```

u=normcdf(d1) ;%permet de calculer la fonction de probabilit\'{e}
    %cumulative de la loi normale centr\'{e}e et r\'{e}duite en d1
v=normcdf(d2) ;

oachat=S*u-X*exp(-r*time)*v; ovente= X*exp(-r*time)*(1-v)-S*(1-u);

if indice == 1
    y=oachat;
else
    y=ovente;
end

%%%%%%%%%%%%%
% Cette fonction calcule la volatilit\'{e} implicite selon le mod\ele
% de Black%Scholes calcul\'{e}e par it\'{e}ration.
% Saimi Nabil Copyright2001
% Universit\'{e} du Quebec a Trois-Rivieres
%%%%%%%%%%%%%

function y=implicitvol(S,X,time,r,targetprice,indice)

high=1.5; low=0; while (high-low)>0.0001

if BS(S,X,(high+low)/2,time,r,indice)>targetprice
    high=(high+low)/2;
else
    low=(high+low)/2;
end
end y=(high+low)/2;

```

Références

- [1] Y. Ait-Sahalia. "Nonparametric Pricing of Interest Rate Derivative Securities." *Econometrica* 64 (1996): 527-60.
- [2] R. G. Avesani et P. Bertrand. *Does volatility jump or just diffuse? a statistical approach*. In L. C. G. Rogers and D. Talay, editors, *Numerical Methods in Finance*, pages 270-289. Cambridge University Press, 1994.
- [3] F. Black et M. Scholes. *The pricing of options and corporate liabilities*. *Journal of Political Economy*, 81:637-654, 1973.
- [4] T. Bollerslev. *Generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics*, 31:309-328, 1986.
- [5] M. Chesney et L. Scott. *Pricing european currency option: A comparison of modified Black&Scholes and a random variance model*. *Journal of Financial and Quantitative Analysis*, 24(3):267-284, 1989.
- [6] Christie, A. A. The stochastic behaviour of common stock variance: value, leverage and interest rate effects. *Journal of Financial Econometrics*, 10, 1982, p407-432.

- [7] Cox, J.C., and Ross, S.A., The Valuation of Options for Alternative Stochastic Processes, *Journal of Financial Economics* 3 (1976), 145-166.
- [8] Droesbeke, J.J. Fichet, B. Tassi, P(1994) *Modélisation ARCH, théorie statistique et application dans le domaine de la finance*. Editions de l'université de Bruxelles, 242 pages.
- [9] J.-C. Duan, Augmented GARCH(p,q) process and its diffusion limit, *Journal of Econometrics* Vol.79(1997)p.97-127.
- [10] D. Duffie. *Dynamic Asset Pricing Theory, Second Edition..* Princeton University Press, Princeton, New Jersey, 1996.
- [11] D. Duffie and K. Singleton. Simulated moments estimation ok Markov models of asset prices. *Econometrica*, 61, 1993, p929-952
- [12] R. Durrett. *Stochastic calculus, A Practical Introduction.* CRC Press, New-York.
- [13] L. Eisenberg. Random variance option pricing and spread valuation. University of Pennsylvania, 1985.
- [14] R. F. Engle and G. G. J. Lee. Estimating diffusion models of stochastic volatility. In P. E. Rossi, editor, *Modeling Stock Market Volatility-Bridging the Gap to Continuous Time*, pages 333-355. Academic Press, 1996.

- [15] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987-1007, July 1982.
- [16] R. F. Engle. *ARCH selected reading*. Cambridge university Press, Cambridge, 1995.
- [17] Fouque, J-P, Papanicolaou G., Sircar K. R, *Derivatives in Financial Markets with Stochastic Volatility*, Editions Cambridge University Press, 2000, 201 pages, ISBN 0 521 79163 4.
- [18] C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8:S85-S118, 1993.
- [19] R. Geske. The valuation of compound options. *Journal of Financial Economics*, 7:63-81, 1979.
- [20] A.C Harvey, *Forecasting, Structural times series models and the Kalman filter*, Cambridge university press, 1989.
- [21] J. Hull, A. White. *The pricing of options on Assets with Stochastic Volatilities*. The Journal of Finance, Vol 42, 1987, pp281-300.
- [22] S. Herzel, *Option pricing with Stochastic Volatility models*. Decision in Economics and Finance, Vol 23, 2000, pp75-99.
- [23] M. Harrison and S. Pliska. Martingales and stochastic integrals in the theroy of continous trading. *Stochastic Process and Their Applications*, 11:215-260, 1981.

- [24] A. C. Harvey. *Forecasting, Structural Models and the Kalman Filter*. Cambridge University Press, New York, 1989.
- [25] A. H. Jaswinski. *Stochastic Process and Filtering Theory*. Academic Press, New-York, 1970, 366 pages.
- [26] H. Johson and D. Shanno. Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis*, 22(2):143-151, 1987.
- [27] C. M. Jarque et A. K. Bera, Efficient tests for normality, heteroscedasticity and a serial indepedence of regression residuals, *Economic Letters* Vol.6, 1980, p.29-61.
- [28] I.Karastzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus, Second Edition*. Springer-Verlag, New-York, 1996.
- [29] Lacoste, V.(1996) *Wiener Chaos:A New Approach To Option Hedging. Mathematical Finance* **Vol.6, No.2**(April 1996) pp. 197-213.
- [30] D. Lamberton et B. Lapeyre. *Introduction au Calcul Stochastique Appliqué à la Finance*. Editions Ellipses. 1991. ISSN: 1154-482 X. 168 pages.
- [31] C. Lecourt, *Les variations de taux de change au jour le jour. Une approche économétrique à partir des processus à mémoire longue*. Thèse de Doctorat en sciences économiques, Université des Sciences et Technologies de Lille 1, Janvier 2000, 371 pages.
- [32] Martineau, B (1997) *La prévision de la volatilité implicite des options sur l'indice boursier S&P500*, Travail dirigé présenté en vue de

l'obtention du grade de maîtrise ès sciences (M.Sc.), École des Hautes Études Commerciales de Montréal, 87 pages.

- [33] B.Mandelbrot. The variation of certain speculative prices. *Journal of Business*,36:394-419, 1963.
- [34] P.S Maybeck. *Stochastic Models, Estimation and Control*. Academic Press, London, 1982.
- [35] R.C. Merton, The Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science* 4 (Spring 1973), 141-183.
- [36] R.C. Merton, Option Pricing When Underlying Stock Returns are Discontinuous, *Journal of Financial Economics* 3 (March 1976), 125-144.
- [37] M. Musiela et M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer, 1997.
- [38] Maddala G.S., Rao C.R, *Statistical Methods in Finance, Handbook of Statistics*, Vol14,733 pages.Part III, pages 119-235.
- [39] J. N. Nielsen, M. Vestergaard et H. Madsen, Estimation in Continuous-Time Stochastic Volatility Models Using NonLinear Filters, *International Journal of Theoretical and Applied Finance*, 2000.
- [40] Ouertani, N (1999) *Etude de la structure de la volatilité implicite des options d'achat sur l'indice boursier S&P500*, Travail dirigé présenté en vue de l'obtention du grade de maîtrise ès sciences (M.Sc.), École des Hautes Études Commerciales de Montréal, 87 pages.

- [41] E. Pardoux et D. Talay, Discretisation and Simulation of Stochastic Differential Equations, *Acta Applicandae Mathematicae*, 3 (1985), p. 23-47.
- [42] Scott, L.O., Option Pricing When the Variance Changes Randomly : Theory, Estimation and an Application, *Journal of Financial and Quantitative Analysis* 22 (1987), 419-438.
- [43] Stein, E.M., and Stein, J.C., Stock Price Distribution with Stochastic Volatility : An Analytic Approach, *Review of Financial Studies* 4 (1991), 727-752.
- [44] Strook, D.W. and S.R.S. Varadhan (1979) *Multi-dimensional Diffusion Process*. Sprind-Verlag, Heidelberg.
- [45] H. Tanizaki. *Nonlinear Filters, Second Edition*. Springer Verlag, Heidelberg, 1996.
- [46] M. Vestergaard, *Nonlinear filtering methods for stochastic volatility models*, Master's Thesis, Dept. of Math. Modelling, Lyngby, Denmark (1999)
- [47] Wan, E.A. Merwe, R.V.D(1997) *The Unscented Kalman Filter for Non-linear Estimation*.
- [48] Watanabe, T.(1999) *A Non-linear filtering approach to stochastic volatility models with an application to daily stock returns*.*Journal of Applied Econometrics* 14 pp. 101-121.
- [49] Welch, G. Bishop, G.(1997) *An Introduction to the Kalman Filter*.

- [50] Wiggins, J.B, Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19:351-372, 1987.