

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATION DE MODÈLES À VOLATILITÉ STOCHASTIQUE AVEC
APPLICATIONS DANS LA COUVERTURE DE FONDS DISTINCTS

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR
JONATHAN GRÉGOIRE

JUIN 2016

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

Remerciements

Je remercie mon superviseur de mémoire Mathieu Boudreault qui est professeur au Département de mathématiques à l'UQAM. Tout au long du processus, il a démontré une grande implication dans ce projet. Il m'a permis d'acquérir des notions très approfondies lors de nos nombreuses rencontres et discussions. Je suis très impressionné par toutes ses réalisations et ses qualités (rigueur, générosité, attention, dévouement, etc). Il est sans aucun doute un exemple à suivre. Sans son soutien, sa confiance et son transfert de connaissances, ma réussite n'aurait pas été envisageable.

Je tiens également à remercier l'Autorité des marchés financiers (AMF) pour son soutien financier. Également, cette organisation m'a permis d'effectuer un stage pratique dans le contexte des fonds distincts. En plus de cette opportunité, mon superviseur de stage, Emmanuel Hamel, m'a permis d'acquérir des connaissances pratiques utiles pour les applications du mémoire.

Sur le plan personnel, il y a beaucoup de gens qui ont cru en moi. Je tiens à manifester ma gratitude envers mes parents (Manon et Jean-François) qui m'ont donné un support inconditionnel me permettant de réaliser tout ce dont j'avais envie. Ma partenaire de vie, Carol-Anne, a nécessairement été LA ressource indispensable lors de cette aventure.

Une dernière reconnaissance pour Maciej Augustyniak pour son aide quant à l'estimation du modèle MS-GARCH et pour ses autres conseils.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vii
LISTE DES FIGURES	ix
RÉSUMÉ	xi
INTRODUCTION	1
CHAPITRE I	
ESTIMATION DU MODÈLE À VOLATILITÉ STOCHASTIQUE EN TEMPS DISCRET	7
1.1 Le modèle SVOL	8
1.1.1 Autres cas particuliers dans la littérature	9
1.2 Synthèse des procédures d'estimation dans la littérature	14
1.2.1 Estimation basée sur une linéarisation du modèle	15
1.2.2 Estimation basée sur le vrai modèle	17
1.3 Méthodes d'estimation utilisées dans ce mémoire	19
1.3.1 Filtre non-linéaire discret (DNF)	21
1.3.2 Filtre à particules (PF)	26
1.4 Étude Monte Carlo	33
1.4.1 Étude de la fonction de vraisemblance	34
1.4.2 Étude Monte Carlo sur les propriétés statistiques du DNF	40
1.4.3 Conclusion	43
1.5 Exemple empirique	43
1.5.1 Comparaison entre des variantes du SVOL	44
1.5.2 Comparaison avec des modèles financiers populaires	47
CHAPITRE II	
ESTIMATION DU MODÈLE À VOLATILITÉ STOCHASTIQUE EN TEMPS CONTINU (HESTON)	51

2.1	Le modèle à volatilité stochastique (temps continu)	52
2.1.1	Le modèle de Heston sous la mesure de probabilité \mathbb{Q}	53
2.2	Synthèse des techniques d'estimation	54
2.2.1	Calibration	55
2.2.2	Estimation statistique	58
2.3	Simulation du modèle de Heston	63
2.3.1	Discrétisation selon les schémas d'Euler	65
2.3.2	Discrétisation selon les schémas de Milstein	66
2.3.3	La méthode exacte	67
2.4	Méthode proposée pour l'estimation du modèle de Heston	67
2.4.1	Quelques ajustements aux méthodes proposées	69
2.5	Étude Monte Carlo	70
2.5.1	Étude de la fonction de vraisemblance	70
2.5.2	Étude Monte Carlo sur les propriétés statistiques du DNF	75
2.5.3	Conclusion	76
2.6	Exemple empirique	77
CHAPITRE III		
EFFICACITÉ DE STRATÉGIES DE COUVERTURE		81
3.1	Construction du marché dans un fonds distinct	82
3.1.1	Dynamique du sous-jacent	82
3.1.2	Dynamique de la valeur du fonds d'investissement	82
3.1.3	Autres hypothèses de marchés	83
3.1.4	Le double risque du marché	84
3.2	Les prestations garanties offertes par l'assureur	84
3.2.1	Prestation minimale garantie à l'échéance ou au décès	84
3.2.2	Dynamique de la valeur garantie G_t	85
3.3	Efficacité des stratégies de couverture	87

3.3.1	Modèle sous la mesure de probabilité neutre au risque \mathbb{Q} . . .	88
3.3.2	Couverture delta d'une garantie minimale à l'échéance (GMMB) avec une prime unique	89
3.3.3	Résultats numériques	93
	CONCLUSION	97
	APPENDICE A	
	RÉSULTATS DU MAXIMUM DE VRAISEMBLANCE	99
	RÉFÉRENCES	103

LISTE DES TABLEAUX

Tableau	Page	
1.1	Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) selon les 50 000 jeux de paramètres et trajectoires aléatoires.	38
1.2	Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) des quatre groupes en fonction des coefficients de variation selon les intervalles de l'équation (1.50).	39
1.3	Résultats numériques de l'étude de la convergence du DNF (1ère partie : $\rho = 0$ et $\rho = -0.3$) où le biais moyen est répertorié ainsi que l'écart-type entre parenthèse.	41
1.4	Résultats numériques de l'étude de la convergence du DNF (2e partie : $\rho = -0.5$ et $\rho = -0.7$) où le biais moyen est répertorié ainsi que l'écart-type entre parenthèse.	42
1.5	Résultats du maximum de vraisemblance pour les trois modèles SVOL ayant des versions différentes selon l'indice composé du S&P/TSX.	45
1.6	Résultats du maximum de vraisemblance pour les trois modèles SVOL ayant des versions différentes selon l'indice composé du S&P500.	46
1.7	Résultats du maximum de vraisemblance pour les cinq modèles stochastiques selon les indices composés du S&P/TSX et S&P500. Pour les détails sur les paramètres, se référer au tableau A.1 dans l'annexe.	49
2.1	Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) selon les 50 000 jeux de paramètres et trajectoires aléatoires.	74
2.2	Résultats numériques de l'étude des propriétés asymptotiques du DNF dans le modèle de Heston. Le biais sur les paramètres et l'écart-type entre parenthèses y sont inscrit.	77
2.3	Résultats du maximum de vraisemblance pour les trois modèles SVOL selon les indices composés du S&P/TSX et S&P500	79

3.1	Résultats de la perte non-couverte de la garantie GMMB. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.	93
3.2	Résultats des erreurs de réplifications $Err_{(T)}$ de la garantie GMMB où la fréquence de re-balancement est mensuelle. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.	94
3.3	Résultats des erreurs de réplifications $Err_{(T)}$ de la garantie GMMB où la fréquence de re-balancement est hebdomadaire. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.	94
3.4	Résultats des erreurs de réplifications $Err_{(T)}$ de la garantie GMMB où la fréquence de re-balancement est journalière. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.	95
3.5	La prime unique π et la volatilité implicite utilisée dans le modèle de Black-Scholes sous la mesure \mathbb{Q} lors des résultats des erreurs de couverture du GMMB.	95
A.1	Résultats du maximum de vraisemblance pour les quatre modèles stochastiques selon les indices composés du S&P/TSX et S&P500.	100
A.2	Résultats du maximum de vraisemblance pour le modèle MS-GARCH selon les indices composés du S&P/TSX et S&P500.	101
A.3	Résultats du maximum de vraisemblance pour les cinq modèles stochastiques selon les indices composés du S&P/TSX et S&P500.	101

LISTE DES FIGURES

Figure		Page
1.1	Histogramme et graphique des log-vraisemblances des deux méthodes DNF et PF selon 50 000 jeux de paramètres et trajectoires aléatoires.	36
1.2	Graphique des erreurs relatives avec deux échelles d'axe en « y » différents selon les 50 000 jeux de paramètres et trajectoires aléatoires.	37
2.1	Histogramme et graphique des log-vraisemblances des deux méthodes DNF et PF selon 50 000 jeux de paramètres et trajectoires aléatoires.	72
2.2	Graphique des erreurs relatives avec deux échelles d'axe en « y » différents selon les 50 000 jeux de paramètres et trajectoires aléatoires.	73
2.3	Graphique des erreurs relatives en fonction de la condition de Feller selon les 50 000 jeux de paramètres et trajectoires aléatoires. . . .	75

RÉSUMÉ

Le modèle à volatilité stochastique, qu'il soit défini en temps discret (comme Taylor (1986)) ou en temps continu (comme Heston (1993)), est un modèle très populaire pour représenter l'évolution temporelle du prix d'actifs financiers. Cette popularité provient du fait que ce modèle incorpore plusieurs faits stylisés et permet de générer des rendements asymétriques avec des queues de distribution plus lourdes que le modèle de Black-Scholes. Toutefois, l'estimation par maximum de vraisemblance du modèle à l'aide d'une seule série temporelle financière est compliquée par le fait que la volatilité n'est pas observable. Ce mémoire analyse numériquement la méthode de filtrage par chaîne de Markov cachée proposée par MacDonald & Zucchini (1997) sur le modèle à volatilité stochastique à temps discret et propose une extension pour estimer le modèle de Heston à temps continu. Cette dernière sera également étudiée numériquement. Pour ces deux modèles, nous comparerons l'approximation par chaîne de Markov cachée avec les filtres à particules. Dans une étude empirique réalisée sur les rendements d'actifs financiers, l'adéquation des modèles à volatilité stochastique est supérieure à plusieurs modèles populaires. Par conséquent, ces modèles sont des alternatives intéressantes pour les assureurs modélisant des fonds distincts en comparaison à d'autres modèles souvent utilisés dans la pratique (par exemple : log-normal à changements de régimes (RSLN)). Finalement, en utilisant les paramètres trouvés précédemment et en comparant avec des modèles populaires, ce mémoire analyse l'efficacité de la stratégie de couverture delta neutre déterminé avec Black-Scholes lorsque le modèle de marché est inconnu. L'idée de l'application est d'analyser la taille de l'erreur de couverture dans un contexte de gestion des fonds distincts venant à échéance à très long terme.

Mots clés : volatilité stochastique, Heston, fonds distincts, stratégie de couverture, méthode de filtrage, maximum de vraisemblance.

INTRODUCTION

Les fonds distincts sont des produits jumelant les caractéristiques d'un fonds commun de placement et d'une assurance vie. Dans un fonds commun de placement, l'investisseur paie une prime périodique en échange d'une gestion efficiente de son fonds d'investissement. Un fonds distinct permet également à un assuré d'investir dans les marchés financiers comme le fonds commun de placement tout en ayant une forme de protection quelconque lors des rendements défavorables du marché.

La protection prend la forme soit d'une rente périodique minimale ou d'une garantie sur la valeur du fonds à l'échéance effective. Comme l'assurance vie, les protections s'appliquent lors d'événements en lien avec le décès ou la survie de l'assuré. Ce produit incorpore donc des caractéristiques actuarielles ainsi que financières.

Les fonds distincts sont une alternative visant à compléter les produits conventionnels d'assurance vie. Plusieurs avenants aux contrats permettent d'avoir un produit attrayant se distinguant d'un assureur à l'autre. Cela permet aux principaux intéressés de s'exposer à long terme aux rendements des marchés financiers et de planifier leur retraite en offrant plus de flexibilité. Par conséquent, les personnes visées par les assureurs sont les gens avers au risque de marché approchant la retraite (âgés de 40 à 55 ans).

Généralement, l'assuré choisit le gestionnaire de son fonds en fonction de son profil de risque (i.e. : un fonds en capital actions, obligations, diversifié, etc.). Par

la suite, l'assureur s'engage à remplir les obligations dictées au contrat. Pour le reste du document, il sera supposé que l'assuré choisit d'investir tout son capital dans un fonds d'investissement qui réplique parfaitement les rendements des indices du S&P/TSX ou du S&P500.

Dans l'assurance traditionnelle, la mortalité est le principal et unique risque. Ce risque est généralement supposé comme étant diversifiable car les compagnies d'assurance ont suffisamment d'assurés. Par conséquent, la théorie des grands nombres s'applique. Donc, les hypothèses actuarielles vont être quantifiées comme étant déterministes.

Contrairement à l'assurance traditionnelle, les fonds distincts font face au risque de marché qui est un risque non-diversifiable. Lors de rendements défavorables sur les marchés, tous les contrats vont être affectés négativement. Ainsi, les compagnies d'assurance doivent élaborer une gestion active de ce risque. L'idée est de créer des stratégies pour répliquer les garanties offertes. Donc, ils vont prendre des positions dans des instruments transigés sur le marché de sorte à contrebalancer leurs obligations.

Pour comprendre comment gérer le passif actuariel lié aux fonds distincts, il faut en connaître la distribution des résultats. Il est donc nécessaire d'élaborer un générateur de scénarios économiques. Ainsi, l'assureur pourra évaluer la différence entre son passif et son actif sur chacun des scénarios en supposant une stratégie de couverture. Cela va permettre de fixer la prime du contrat en fonction du risque résiduel et de la profitabilité visée.

À ce jour, il existe plusieurs modèles réalistes pour générer les prix d'un indice financier mais aucun qui ne soit parfaitement valable. Généralement, les modèles

les plus sophistiqués améliorent la qualité de l'adéquation mais sont également plus difficiles à estimer. Ainsi, l'actuaire doit choisir entre un modèle bien estimé à l'adéquation insuffisante ou un modèle sophistiqué mal estimé.

La création de modèles financiers se base principalement sur l'observation de faits stylisés des rendements financiers. Par exemple, une volatilité stochastique, des changements de régimes, des sauts, un effet de levier entre la volatilité sont des exemples de faits stylisés qui ont influencé certains auteurs à proposer un nouveau modèle dans la littérature financière. Le but est de représenter les rendements boursiers avec un modèle suffisamment sophistiqué captant plusieurs faits stylisés.

Un modèle populaire au Canada pour la tarification des fonds distincts est le modèle lognormal à changements de régimes (RSLN). Une alternative intéressante est la classe de modèles incorporant une volatilité stochastique et latente. Parmi cette classe, on retrouve le modèle à volatilité stochastique à temps discret (SVOL) et le modèle de Heston à temps continu. Également, dans un rapport publié en 2005 par l'American Academy of Actuaries (AAA), ce groupe recommande fortement l'utilisation d'un modèle à volatilité stochastique pour la tarification des fonds distincts. Cependant, tous ces modèles nécessitent des méthodes numériques exigeantes en temps de calcul pour en estimer les paramètres. Le principal objectif de ce mémoire est d'analyser l'estimation par maximum de vraisemblance de divers modèles à volatilité stochastique avec l'objectif que ces modèles puissent être utilisés dans la gestion des risques liés aux fonds distincts.

Après avoir fait une revue exhaustive de la littérature, nous appliquerons tout d'abord la méthode des chaînes de Markov cachée de Zucchini & McDonald (1997) et Clements et al. (2004) à l'estimation de différents modèles à volatilité stochastique en temps discret. Ensuite, nous proposerons une extension de la méthode

pour estimer le modèle à temps continu de Heston. Au meilleur de nos connaissances, il n'existe pas de méthode efficiente d'estimation du modèle de Heston basée sur le maximum de vraisemblance qui utilise seulement les prix d'une action. Finalement, nous présenterons une brève application à l'analyse de l'efficacité de la couverture de fonds distincts. La structure du mémoire est présentée ci-bas.

Premièrement, le chapitre 1 présente une revue de la littérature concernant les modèles SVOL à temps discret ainsi que les différentes méthodes d'estimation. La section 1.3 approfondit deux techniques d'estimation soit les chaînes de Markov cachées et les filtres à particules. Par la suite, la section 1.4 sert à vérifier le comportement numérique de ces deux techniques d'estimation pour valider leur utilisation. Finalement, la section 1.5 permet de comparer l'adéquation de modèles à volatilité stochastique par rapport à certains autres modèles populaires de la littérature.

Deuxièmement, la construction du chapitre 2 est très similaire au chapitre 1 mais porte sur le modèle de Heston à temps continu. Encore une fois, une revue de la littérature concernant les méthodes d'estimation de ce modèle y est présentée. Les deux mêmes techniques présentées plus tôt y sont appliquées. Les sections 2.3 et 2.4 proposent l'adaptation nécessaire pour permettre l'estimation de ce modèle. Dans un premier temps, il y aura une discrétisation du modèle continu vers un modèle à temps discret. Ensuite, nous proposerons des modifications à la technique d'estimation pour qu'elle demeure valide. La section 2.5 est similaire à la section 1.4 et permet l'analyse de l'efficacité des techniques d'estimation avec le modèle de Heston. La section 2.6 est une extension de la section 1.5 où la comparaison est étendue en incluant le modèle de Heston.

Le chapitre 3 sert à étudier un contexte actuariel réaliste et analyse le risque

de modèle dans l'efficacité de stratégies de couverture pour la gestion de fonds distincts. Tous les modèles présentés précédemment sont estimés sur deux bases de données (S&P500 et S&P/TSX). Dans la section 3.3, la stratégie de couverture delta neutre dans le modèle de Black-Scholes est analysée lorsque le modèle de marché y est inconnu (SVOL, Heston, RSLN, etc). Ce chapitre est davantage une application numérique simple pour recréer les risques attachés aux fonds distincts.

CHAPITRE I

ESTIMATION DU MODÈLE À VOLATILITÉ STOCHASTIQUE EN TEMPS DISCRET

1.1 Le modèle SVOL

Le modèle à volatilité stochastique (SVOL) à temps discret a été introduit par Taylor (1986). Plusieurs formes sont apparues dans la littérature financière et peuvent être pratiquement toutes écrites comme :

$$\begin{aligned} y_t &= \lambda g(y_{t-1}, x_t, x_{t-1}) + \sigma_t \epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \delta p(y_{t-1}, x_{t-1}) + \sigma_\nu \nu_t \end{aligned} \tag{1.1}$$

où $\sigma_t = \exp(x_t/2)$ et $\{\gamma, \phi, \sigma_\nu, \lambda, \delta\}$ sont des paramètres. Dans cette représentation, les log-rendements y_t sont considérés comme étant observés¹ tandis que la log-variance x_t est supposée latente, non-observée. Une hypothèse simpliste serait de considérer ϵ_t et ν_t comme étant des bruits gaussiens indépendants. Cependant, dans le but de représenter l'effet de levier observé sur les marchés financiers, il vaut mieux considérer ϵ_t et ν_t comme des variables corrélées mais la paire (ϵ_t, ν_t) comme étant i.i.d. dans le temps. Cela est possible lorsque $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$. Il est également possible d'y inclure une certaine « corrélation » en incluant un « *threshold* » dans l'équation de x_t avec la fonction $\delta p(y_{t-1}, x_{t-1})$ comme dans les modèles classiques EGARCH, NGARCH, GJR-GARCH². Par exemple, en lien avec le modèle EGARCH, il serait possible de supposer $\delta p(y_{t-1}, x_{t-1}) = \delta_1(|\epsilon_{t-1}| - \mathbb{E}[|\epsilon_{t-1}|]) + \delta_2 \epsilon_{t-1}$ qui captera en quelque sorte l'asymétrie des rendements.

1. Le log-rendement y_t peut être obtenu comme $y_t = \log\left(\frac{S_{t+1}}{S_t}\right)$ où S_t est le prix d'un actif financier.

2. Bollerslev (1986) a introduit le modèle GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*). Plusieurs extensions de ce modèle ont été proposées par la suite comme les modèles EGARCH (*exponential GARCH*) par Nelson (1991), le GJR-GARCH par Glosten, Jagannathan et Runkle (1993) puis le NGARCH (non-linear GARCH) par Engle et Ng (1993).

La fonction $\lambda g(y_{t-1}, x_t, x_{t-1})$ sert à spécifier la dérive du processus de y_t . Le cas le plus simple serait tout simplement de poser $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda = \mu$, une constante fixant la moyenne des rendements. Toutefois, la fonction permet d'être un peu plus flexible. Comme le modèle « GARCH in mean, GARCH-M », elle permet une portion auto-régressive sur les rendements ainsi que l'ajout d'un terme hétéroscédastique en fonction de $\sigma_t^2 = \exp(x_t)$. Alors, un autre choix serait de poser $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda_1 + \lambda_2 y_{t-1} + \lambda_3 \sigma_t^2$. Dans la littérature, l'estimation du modèle se fait généralement en fixant $\lambda g(y_{t-1}, x_t, x_{t-1}) = 0$ tout en utilisant une base de données où la moyenne est soustraite ainsi que les autres composantes auto-régressives et saisonnières.

1.1.1 Autres cas particuliers dans la littérature

1.1.1.1 Version la plus répandue

Plusieurs cas particuliers du modèle (1.1) sont apparus dans la littérature. Par exemple, la version la plus simple s'écrivant :

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \tag{1.2}$$

où $\mathbb{E}[\epsilon_t \cdot \nu_t] = \lambda g(y_{t-1}, x_t, x_{t-1}) = \delta p(y_{t-1}, x_{t-1}) = 0$. En supposant des perturbations gaussiennes, les travaux de Melino & Turnbull (1990), Duffie & Singleton (1993), Jacquier et al. (1994), Danielsson & Richard (1993) et Shephard (1993) ont investigué ce modèle. Il y a eu également les travaux de Nelson (1988), Harvey et al. (1994), Kim et al. (1998), Singleton (2001) et Knight et al. (2002) mais dans une version linéarisée (voir l'équation (1.8)).

1.1.1.2 Version avec des bruits non-gaussiens

Certains auteurs ont permis l'extension du modèle avec des bruits ayant des queues plus lourdes dans l'équation d'observation. Ce cas ne sera pas analysé en détail

puisque le but ultimement de ce mémoire est d'utiliser une transformation du processus de y_t (i.e. : S_t). Très souvent, dans les contrats de type fonds distinct, il faut calculer des valeurs d'options (ou primes) mais également des mesures de risques (VaR, TVaR, ...) pour déduire ces primes. Une condition nécessaire pour avoir des prix d'options $C(t, S_t)$ avec une variance finie (i.e. : carré intégrable), il est requis que

$$\mathbb{E}[S_t^2] < \infty.$$

Donc, il est préférable d'avoir des perturbations venant de lois gaussiennes, car l'exponentielle d'une Student, par exemple, ne sera pas toujours de carré intégrable.

1.1.1.3 Extensions pour une meilleure adéquation

D'autres variantes du modèle de l'équation (1.2) sont apparues pour mieux capter la moyenne de y_t et l'adéquation de la variable latente x_t . Geweke (1994) propose d'inclure des variables exogènes dans l'équation de y_t qui se résume à poser $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda' z_t$ où z_t est un vecteur de variables exogènes et λ' un vecteur de paramètres. Fridman & Harris (1998) ont proposé quelques variantes du modèle dont la plus générale avec un GARCH dans la moyenne et un ajustement dans l'adéquation de la variable latente se résumant à poser $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda \sigma_t^2$ et $\delta p(y_{t-1}, x_{t-1}) = \delta |y_{t-1}|$. Dans la même optique, Watanabe (1999) propose de mieux prédire la moyenne en fixant la dérive par $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda_0 + \lambda_1 y_{t-1} + \lambda_2 y_{t-2} + \lambda_3 \sigma_t^2 + \lambda_4 D_t$ et de rajouter des termes dans l'équation de x_t en fixant $\delta p(y_{t-1}, x_{t-1}) = \delta_1 D_t + \delta_2 \epsilon_{t-1} + \delta_3 |\epsilon_{t-1}|$ où D_t est une variable indicatrice saisonnière (fin de semaine). Évidemment, ces exemples sont seulement quelques cas particuliers parmi ceux proposés dans la littérature financière.

1.1.1.4 Présence de l'effet de levier

Une extension intéressante du modèle SVOL consiste à incorporer un effet de levier entre les rendements et la log-variance. Cependant, les premiers travaux se basaient surtout sur la version transformée (basée sur $\log y_t^2$) et l'information relative à la corrélation entre les processus y_t et x_t est perdue. Pour pallier à ce problème, Harvey et Shephard (1996) ont démontré comment il était possible de retrouver cette information en conditionnant sur le signe (s_t) de y_t . Cela n'altère pas la distribution du bruit du $\log \epsilon_t^2$ et il en résulte ce modèle :

$$\begin{aligned} \log y_t^2 &= x_t + \log \epsilon_t^2 \\ x_t &= \phi x_{t-1} + A s_t + \tilde{\nu}_t, \quad \tilde{\nu}_t \sim IID(0, \sigma_\nu^2 - A^2) \end{aligned} \tag{1.3}$$

où $A = 0.7979\rho\sigma_\nu$, $B = 1.1061\rho\sigma_\nu$, et $\mathbb{E}[\log \epsilon_t^2 \cdot \nu_t] = B s_t$. Sachant que ϵ_t et ν_t sont des normales bivariées où $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$, il est possible d'en déduire que $\mathbb{E}[\nu_t | \epsilon_t] = \rho\sigma_\nu\epsilon_t$. En dénotant l'espérance conditionnelle sachant que ϵ_t est positif par \mathbb{E}^+ , les constantes A et B proviennent des relations suivantes :

$$\begin{aligned} A &= \mathbb{E}^+[\nu_t] = \rho\sigma_\nu\mathbb{E}^+[\epsilon_t] = \rho\sigma_\nu\sqrt{2/\pi} \\ B &= \text{cov}^+[\nu_t \cdot \log \epsilon_t^2] = \rho\sigma_\nu\mathbb{E}[|\epsilon_t| \log \epsilon_t^2] - A\mathbb{E}[\log \epsilon_t^2]. \end{aligned}$$

Cependant, la nouvelle distribution conditionnellement sur le signe s_t du bruit $\tilde{\nu}_t$ n'est plus normalement distribuée et causera une perte d'efficacité lors de l'estimation du modèle via le filtre de Kalman. En effet, les filtres de Kalman supposent des perturbations gaussiennes. Ce modèle a également été analysé par Sandmann & Koopman (1998) permettant de capter l'effet de levier, mais avec une autre technique d'estimation.

Par la suite, MacDonald et al. (1997) et Jacquier et al. (2004) ont proposé une

méthode d'estimation sur le modèle suivant :

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \quad (1.4)$$

où $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$, $\lambda g(y_{t-1}, x_t, x_{t-1}) = 0$ et $\delta p(y_{t-1}, x_{t-1}) = 0$. La différence parmi les autres modèles est que l'effet de levier est directement modélisé sans avoir recours à des transformations au sein de l'équation de la variable latente.

Également pour capter l'effet de levier, un modèle SVOL avec un «*threshold*» permettant des paramètres variant dans le temps a été proposé par So et al. (2003) s'écrivant :

$$\begin{aligned} y_t &= \mu_{s_{t-1}} + \beta_{s_{t-1}} y_{t-1} + \sigma_t \epsilon_t \\ x_t &= \gamma_{s_{t-1}} + \phi_{s_{t-1}} x_{t-1} + \sigma_{\nu, s_{t-1}} \nu_t \end{aligned} \quad (1.5)$$

où $\mathbb{E}[\epsilon_t \cdot \nu_t] = 0$, s_t est la variable indicatrice $\mathbb{1}_{\{y_t < 0\}}$ et

$$\begin{aligned} \mu_{s_{t-1}} &= \mu_0 + s_{t-1} \mu_1, & \beta_{s_{t-1}} &= \beta_0 + s_{t-1} \beta_1 \\ \gamma_{s_{t-1}} &= \gamma_0 + s_{t-1} \gamma_1, & \phi_{s_{t-1}} &= \phi_0 + s_{t-1} \phi_1 \\ \sigma_{\nu, s_{t-1}}^2 &= \sigma_{\nu, 0}^2 + s_{t-1} \sigma_{\nu, 1}^2. \end{aligned}$$

Ce modèle peut pratiquement être écrit sous la forme proposé par l'équation (1.1) en fixant $\lambda g(y_{t-1}, x_t, x_{t-1}) = \mu_0 + s_{t-1} \mu_1 + \beta_0 y_{t-1} + s_{t-1} \beta_1 y_{t-1}$ et $\delta p(y_{t-1}, x_{t-1}) = s_{t-1} \gamma_1 + s_{t-1} \phi_1 x_{t-1}$ mais n'incorpore pas la portion reliée au coefficient $\sigma_{\nu, s_{t-1}}^2$. Cependant, les méthodes proposées dans le mémoire pourraient être adaptées avec quelques légères modifications pour prendre en compte cette portion. De plus, Smith (2009) a proposé une extension du modèle (1.5) pour améliorer l'aspect général de l'effet de levier en rajoutant également une corrélation entre les bruits ϵ_t et ν_t .

1.1.1.5 Cas répandu dans la pratique

Un autre modèle espace-état avec une définition légèrement différente pour la variable latente est celui proposé par l'American Academy of Actuaries (AAA, 2005) pour de la modélisation des rendements mensuels ($\Delta_t = 1/12$) dans le contexte des fonds distincts. Ce modèle s'écrit :

$$\begin{aligned} y_t &= (A + B\sigma_t + C\sigma_t^2)\Delta_t + \sqrt{\Delta_t}\sigma_t\epsilon_t \\ \log \sigma_t = x_t &= (1 - \phi)x_{t-1} + \phi \log \tau + \sigma_\nu \nu_t \end{aligned} \quad (1.6)$$

où $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$. Ils présentent ce modèle avec deux restrictions sur la log-volatilité. Ce sont des bornes x^+ et x^- représentant respectivement le maximum et minimum que peut prendre la log-volatilité donc,

$$x^- \leq x_t \leq x^+.$$

À noter que, hormis les deux restrictions, ce modèle pourra être estimé avec les méthodes présentées plus bas en modifiant légèrement les densités conditionnelles de y_t et x_t .

Bref, le modèle proposé permet d'englober un vaste nombre de modèles sans avoir recours à des transformations supplémentaires. Son utilisation peut être appliquée dans plusieurs contextes financiers (taux de change, rendements boursiers, taux d'intérêt, etc.) en incorporant facilement diverses facettes connues de la littérature selon la représentation de y_t . Par exemple, il pourrait être intéressant d'avoir une auto-corrélation sur la valeur absolue des rendements qui est facilement intégrable dans la fonction $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda_1 |y_{t-1}|$. Contrairement à certaines méthodes d'estimation dans la littérature, les méthodes présentées plus bas permettront d'avoir des estimés sur tous les paramètres en une seule maximisation.

1.2 Synthèse des procédures d'estimation dans la littérature

Cette partie analysera principalement les procédures dont l'objet est la maximisation de la fonction de vraisemblance. Dans le cas du modèle SVOL, il est impossible d'écrire cette fonction de manière analytique. Pour une série de données $y_{1:T} = \{y_1, \dots, y_T\}$ et de variables d'état $x_{0:T} = \{x_0, \dots, x_T\}$ où x_0 est la variable d'état initiale, la vraisemblance du modèle SVOL s'écrit :

$$\mathcal{L}(y_{1:T}|\theta) = \int \dots \int f(y_{1:T}|x_{0:T}, \theta) f(x_{0:T}|\theta) dx_0, \dots, dx_T \quad (1.7)$$

où $\theta = \{\gamma, \phi, \sigma_\nu, \lambda, \delta, \rho\}$ est un vecteur de paramètres.

Dans une représentation générale des modèles sous la forme espace-état, il est possible de résoudre ces intégrales avec des filtres de Kalman (KF) si les perturbations du système sont gaussiennes et que le lien entre l'équation d'observation et l'équation de la variable latente est linéaire. Sinon, il faut des filtres de Kalman plus sophistiqués (filtre de Kalman étendu (EKF, *Extended KF*) ou filtre de Kalman inodore (UKF, *Unscented KF*)) lorsque le lien n'est pas linéaire.

Dans le cas du modèle SVOL, les perturbations de l'équation d'observation ne sont pas gaussiennes et le lien entre y_t et x_t n'est pas linéaire. Alors, le problème demeure dans la résolution de cette intégrale de $T+1$ dimensions. Dans un contexte de maximum de vraisemblance, il vaut mieux travailler avec la log-vraisemblance car numériquement la maximisation de la vraisemblance n'est pas envisageable. Cela engendre l'équation suivante

$$\log \mathcal{L}(y_{1:T}|\theta) = \log \left\{ \int \dots \int f(y_{1:T}|x_{0:T}, \theta) f(x_{0:T}|\theta) dx_0, \dots, dx_T \right\}$$

qui sera référencée comme étant la même équation que (1.7).

1.2.1 Estimation basée sur une linéarisation du modèle

Dans la littérature, une des premières approches pour estimer le modèle a été de linéariser le système venant de l'équation (1.1). En prenant le log du carré de y_t , il en résulte un modèle SVOL transformé s'écrivant :

$$\begin{aligned} \log(y_t^2) &= x_t + \log(\epsilon_t^2) \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \tag{1.8}$$

où $\lambda g(y_{t-1}, x_t, x_{t-1}) = \delta p(y_{t-1}, x_{t-1}) = \rho = 0$ et $\log(\epsilon_t^2)$ correspond au log d'une variable aléatoire distribuée selon la loi du chi-carré (χ^2) avec un degré de liberté. Suite à cette transformation, le système (1.8) est linéaire et additif. Cependant, un nouveau problème survient. Les perturbations de l'équation d'observation ne sont plus gaussiennes et ne peuvent être approximées par les méthodes traditionnelles de filtrage.

Nelson (1988) et Harvey et al. (1994) ont proposé de façon indépendante d'approximer la distribution du $\log(\chi^2)$ par une loi normale de moyenne égale à $\psi(1/2) - \log(1/2)$ et de variance $\pi^2/2$ où ψ est la fonction digamma. À partir du système (1.8) et de cette hypothèse, l'équation (1.7) peut être calculé avec un filtre de Kalman traditionnel.

Dans la littérature du SVOL, cette méthode est souvent référencée sous le nom de « Quasi-Maximum Likelihood » (QML). Harvey et al. (1994) et Ruiz (1994) remarquent que l'approximation du bruit non-gaussien par un bruit gaussien est adéquat seulement lorsque σ_ν est grand. En effet, la composante systémique de x_t dans l'équation d'observation doit dominer le terme d'erreur du $\log \epsilon_t^2$.

Une autre approche a été introduite par Kim et al. (1998) pour régler ce problème. Ils ont approximé le bruit non-gaussien par un mélange de variables normales

jumelé à une procédure bayésienne de re-pondération pour corriger l'erreur de linéarisation. Cependant, les méthodes « Monte Carlo Markov Chain » (MCMC) sont reconnues exigeantes en termes de calculs. Ainsi, Sandmann & Koopman (1998) ont développé une technique basée sur la simulation pour approximer la log-vraisemblance du bruit non-gaussien. Cette technique « Monte Carlo likelihood » (MCL) peut être représentée par cette équation générale

$$\log \mathcal{L}(y|\theta) = \log \mathcal{L}_G(y|\theta) + \log \mathbb{E}_G \left[\frac{f_{\text{true}}(\log \mathcal{X}^2|\theta)}{f_G(\log \mathcal{X}^2|\theta)} \right] \quad (1.9)$$

où G signifie la provenance du modèle transformé avec l'approximation gaussienne tandis que « true » provient de la vraie distribution du $\log \mathcal{X}^2$. Or, la log-vraisemblance du modèle peut être vue comme étant celle du modèle QML additionnée d'un terme de correction basé sur la densité d'importance $f_G(\log \mathcal{X}^2|\theta)$. Contrairement à la méthode MCMC, Sandmann & Koopman mentionnent qu'il est possible d'être précis seulement avec cinq simulations ($N = 5$) rendant la procédure plus rapide.

Singleton (2001) et Knight et al. (2002) ont proposé une approche complètement différente basée sur la fonction caractéristique. Le but sera d'égaliser la fonction cumulative théorique du modèle transformé à la fonction caractéristique empirique (ECF). Ainsi, l'estimation reposera sur un critère de minimisation de la distance entre ces deux fonctions.

Ces différentes approches permettent de simplifier la résolution de l'intégrale (1.7) mais rendent plus difficile l'ajout de faits stylisés des marchés financiers (variables exogènes, auto-régression, effet de levier, etc). En effet, il faut souvent considérer $\lambda g(y_{t-1}, x_t, x_{t-1}) = 0$ pour que la transformation soit possible et maintenir un système linéaire. Ainsi, l'estimation du modèle se fera en deux temps. Il faudra estimer de manière empirique les constantes et composantes de saisonnalité au

sein de la fonction $\lambda g(y_{t-1}, x_t, x_{t-1})$. Par la suite, la base de données sera modifiée en enlevant ces composantes précédemment estimées. Finalement, l'estimation des autres paramètres se fera avec une méthode quelconque sur la nouvelle base de données. De plus, l'effet de levier peut être incorporé en utilisant l'astuce de Harvey et Shephard (1996) mais nécessite des hypothèses supplémentaires. En général, cette transformation résout un peu la problématique de l'équation (1.7) mais le modèle devient beaucoup moins flexible.

1.2.2 Estimation basée sur le vrai modèle

D'autres approches d'estimation ont été proposées après Nelson (1988) qui ne reposent pas sur des transformations du modèle. Les deux problèmes de l'équation (1.7) vont être résolus en gardant les propriétés et l'écriture du modèle.

1.2.2.1 Méthodes des moments

Les premiers travaux furent basés sur des méthodes des moments (MM). Cela permet d'ignorer la résolution de la fonction de vraisemblance et de résoudre un système d'équations avec autant d'inconnues. Melino & Turnbull (1990) proposent d'estimer les paramètres via une méthode des moments généralisés (GMM) tandis que Duffie & Singleton (1993) proposent une méthode des moments simulés (SMM). Ces méthodes sont très simples et requièrent peu de calculs informatiques. Cependant, l'efficacité de ces méthodes n'est pas optimale en les comparant à des méthodes bayésiennes ou de maximum de vraisemblance (ML). Tel que démontré dans les articles de Jacquier et al. (1994) et (2004), ces méthodes sont imprécises quant à l'estimation de σ_ν , surtout lorsque l'estimation de $\phi \approx 1$. Les résultats sont encore moins concluant lorsque le coefficient de variation modifié défini comme étant

$$CV = \text{Var}[x_t] / \mathbb{E}[x_t]^2 = \exp(\sigma_\nu^2 / (1 - \phi^2)) - 1 \quad (1.10)$$

est petit. C'est également le cas pour les techniques basées sur QML. Il est important de noter que MM et QML ne sont pas des techniques appropriées car tel que démontré par Jacquier et al. (1994), en économétrie financière le paramètre ϕ est près de 1 et le CV est petit.

1.2.2.2 Maximum de vraisemblance

Plusieurs travaux sur les méthodes numériques de filtrage ont permis l'approximation de l'équation (1.7). En ce sens, Danielsson & Richard (1993) proposent une méthode générale pour estimer les modèles à variables latentes. En prenant N échantillons d'une distribution spécifique $p(\theta)$, l'approximation de l'équation (1.7) devient

$$\hat{\mathcal{L}}(y_1, \dots, y_T | \theta) = \frac{1}{N} \sum_{i=1}^N p^{(i)}(\theta). \quad (1.11)$$

Ainsi, l'application de la méthode du maximum de vraisemblance simulé (SML) a été appliquée au modèle SVOL par Danielsson (1994). Le problème demeure de trouver l'échantillonneur optimal qui permet de bien approximer la fonction $p(\theta)$. Très souvent, cela requière un algorithme d'optimisation exigeant en calcul informatique.

Deux méthodes plus rapides ont été proposées par Fridman & Harris (1998) et Watanabe (1999) toutes les deux basées sur la méthode d'intégration numérique récursive suggérée par Kitagawa (1987). En utilisant les densités conditionnelles de y_t et x_t , il est possible de construire un filtre récursif dans l'optique d'évaluer l'équation (1.7) en calculant successivement plusieurs intégrales par morceaux. La différence est dans la résolution des intégrales par morceaux. Kitawaga (1987) a suggéré de résoudre ces intégrales en approximant certaines densités par plusieurs fonctions linéaires. Fridman & Harris (1998) réclame que la méthode d'intégration Gauss-Legendre est plus efficiente. D'un autre côté, Watanabe (1999) propose de

résoudre ces intégrales en discrétisant la variable latente et en résolvant les intégrales par morceaux avec une méthode trapézoïdale. Ces méthodes ont pour but de maximiser l'équation (1.7) pour approximer l'estimateur de θ .

1.2.2.3 Méthodes bayésiennes

Une approche totalement différente serait d'utiliser des algorithmes bayésiens. Les premiers algorithmes bayésiens ont été proposés par Shephard (1993) et Jacquier et al. (1994) suivis de Geweke (1994). Le but est de créer une chaîne de Markov récursive basée sur une méthode d'échantillonnage qui finira par converger vers la densité *a posteriori*. Ainsi, les paramètres θ seront approximés lorsque la convergence de cette densité sera atteinte. Il existe une très grande littérature au sujet de ces méthodes qui est judicieusement abrégée dans l'article de Jacquier et al. (2004).

En général, les méthodes bayésiennes et de maximum de vraisemblance sont très souvent perçues dans la littérature comme étant équivalentes au niveau computationnelle. De plus, ces deux méthodes sont très souvent exigeantes au niveau informatique en raison des problèmes de convergence pour les méthodes bayésiennes ou en raison de la résolution d'intégrales numériques pour les méthodes maximum de vraisemblance.

1.3 Méthodes d'estimation utilisées dans ce mémoire

Le double objectif est de proposer des méthodes accommodant tous les faits stylisés du modèle (1.1) tout en étant efficace (sans biais et rapide). Deux méthodes seront décrites soit celle du filtre non-linéaire discret (DNF) proposée par MacDonald et al. (1997) et repris par Clement et al. (2004a et 2004b) puis celle d'un filtre à particules (PF).

Ces méthodes requièrent uniquement les densités conditionnelles de y_t et x_t qui serviront essentiellement à appliquer des techniques de filtrage lors de la résolution des intégrales de l'équation (1.7). En appliquant la décomposition de Cholesky et en introduisant un bruit ϵ_t^* indépendant de ν_t , le système (1.1) se réécrit

$$\begin{aligned}\epsilon_t &= \rho\nu_t + \sqrt{1 - \rho^2}\epsilon_t^* \\ \nu_t &= \frac{x_t - \gamma - \phi x_{t-1} - \delta p(y_{t-1}, x_{t-1})}{\sigma_\nu} \\ y_t &= \lambda g(y_{t-1}, x_t, x_{t-1})\Delta_t + \sqrt{\Delta_t} \exp(x_t/2) \left(\rho \frac{x_t - \gamma - \phi x_{t-1} - \delta p(y_{t-1}, x_{t-1})}{\sigma_\nu} + \sqrt{1 - \rho^2}\epsilon_t^* \right) \\ x_t &= \gamma + \phi x_{t-1} + \delta p(y_{t-1}, x_{t-1}) + \sigma_\nu \nu_t\end{aligned}\tag{1.12}$$

où $\mathbb{E}[\nu_t \cdot \epsilon_t^*] = 0$. Cela engendre les distributions conditionnelles suivantes :

$$\begin{aligned}r(y_t|x_t, x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi \exp(x_t)(1 - \rho^2)\Delta_t}} \exp\left(-\frac{(y_t - \mu_t)^2}{2 \exp(x_t)(1 - \rho^2)\Delta_t}\right) \\ q(x_t|x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left(-\frac{[x_t - \{\gamma + \phi x_{t-1} + \delta p(y_{t-1}, x_{t-1})\}]^2}{2\sigma_\nu^2}\right)\end{aligned}\tag{1.13}$$

où Δ_t représente la fréquence d'observation annuelle et

$$\mu_t = \lambda g(y_{t-1}, x_t, x_{t-1})\Delta_t + \sqrt{\Delta_t} \frac{\exp(x_t/2) \rho (x_t - \gamma - \phi x_{t-1} - \delta p(y_{t-1}, x_{t-1}))}{\sigma_\nu}.$$

Ces distributions sont déduites avec l'hypothèse que ϵ_t et ν_t sont des bruits gaussiens mais il est aussi possible de faire l'extension avec des bruits ayant des queues plus lourdes.

Les mêmes démarches peuvent être effectuées pour accommoder certaines variantes de processus à volatilité stochastique dont celle du modèle de AAA référée par l'équation (1.6). Pour ce modèle, les densités conditionnelles de y_t et x_t en ne

considérant pas les restrictions sont :

$$\begin{aligned} r(y_t|x_t, x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi \exp(2x_t)(1-\rho^2)\Delta_t}} \exp\left(-\frac{(y_t - \mu_t)^2}{2 \exp(2x_t)(1-\rho^2)\Delta_t}\right) \\ q(x_t|x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left(-\frac{[x_t - \{(1-\phi)x_{t-1} + \phi \log \tau\}]^2}{2\sigma_\nu^2}\right) \end{aligned} \quad (1.14)$$

où

$$\mu_t = [A + B \exp(x_t) + C \exp(2x_t)]\Delta_t + \sqrt{\Delta_t} \frac{\exp(x_t)\rho(x_t - (1-\phi)x_{t-1} - \phi \log \tau)}{\sigma_\nu}.$$

Dans les deux prochaines sous-sections, les méthodes DNF et PF seront présentées et feront référence à l'écriture de ces deux densités conditionnelles.

1.3.1 Filtre non-linéaire discret (DNF)

Cette méthode a été proposée par MacDonald et al. (1997) et elle permet d'obtenir les paramètres du modèle (1.1) en maximisant l'équation (1.7). Le principal avantage de cette méthode est qu'elle permet d'évaluer directement les intégrales (intégrales doubles lorsqu'on inclut l'effet de levier) dans les filtres non-linéaires pour différentes dynamiques de processus à variables latentes. Ceci est possible en discrétisant le domaine de la variable latente en N intervalles adjacents et en supposant des transitions similaires à une chaîne de Markov. En opposition, Kitawaga (1987), Fridman et al. (1998) et Watanabe (1999) utilisent des techniques d'intégrations différentes pour résoudre le filtre non-linéaire récursif.

1.3.1.1 Trois étapes du filtre non-linéaire

Ces étapes sont grandement inspirées du filtre de Kitawaga (1987) avec quelques modifications. En effet, lors de l'introduction de l'effet de levier, la distribution marginale de y_t ne dépend plus seulement de x_t mais également de x_{t-1} et nécessitera une intégration supplémentaire par rapport à x_{t-1} . Les étapes feront référence

aux densités écrites à l'équation (1.13).

L'étape de prédiction devient :

$$f(x_t, x_{t-1} | y_{t-1}, \theta) = q(x_t | x_{t-1}, y_{t-1}, \theta) f(x_{t-1} | y_{t-1}, \theta) \quad (1.15)$$

et l'étape de mise à jour de x_t conditionnellement à l'information au temps t est donnée par :

$$\begin{aligned} f(x_t | y_t, \theta) &= \int_{-\infty}^{\infty} f(x_t, x_{t-1} | y_t, \theta) dx_{t-1} \\ &= \frac{\int_{-\infty}^{\infty} f(x_t, x_{t-1}, y_t | y_{t-1}, \theta) dx_{t-1}}{f(y_t | y_{t-1}, \theta)} \\ &= \frac{\int_{-\infty}^{\infty} r(y_t | x_t, x_{t-1}, y_{t-1}, \theta) f(x_t, x_{t-1} | y_{t-1}, \theta) dx_{t-1}}{f(y_t | y_{t-1}, \theta)}. \end{aligned} \quad (1.16)$$

Or, il ne reste plus qu'à déduire la fonction de vraisemblance de y_t qui s'écrit :

$$\begin{aligned} f(y_t | y_{t-1}, \theta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_t, x_t, x_{t-1} | y_{t-1}, \theta) dx_t dx_{t-1} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(y_t | x_t, x_{t-1}, y_{t-1}, \theta) f(x_t, x_{t-1} | y_{t-1}, \theta) dx_t dx_{t-1}. \end{aligned} \quad (1.17)$$

L'utilité de la méthode DNF est d'approximer numériquement l'intégrale double de l'équation (1.17) via une technique simple. Ceci va être possible en discrétisant la variable latente et en l'évaluant de la même manière qu'un processus markovien (HMM) d'une série temporelle discrète. L'adaptation du filtre de Smith (2009) est très similaire à celle décrite plus haut sauf que la méthode d'intégration utilisée va dans le sens de celle proposée par Fridman et al. (1998).

1.3.1.2 Discrétisation de la variable latente

L'évaluation du filtre repose sur la discrétisation de la variable latente. Notons \mathbf{X} le vecteur représentant les N points uniformément dispersés sur l'intervalle

suisant :

$$\begin{aligned}\mathbf{X} &= [x^{(1)}, \dots, x^{(N)}] \\ x^{(1)} &= E_0 - I_x, \quad x^{(N)} = E_0 + I_x \\ I_x &= \delta_x(N) \sqrt{V_0}\end{aligned}\tag{1.18}$$

où E_0 et V_0 signifient respectivement l'espérance et la variance du processus stationnaire de la variable latente³. La distance Δ_x engendrée entre chaque points $x^{(k)}$ sera de

$$\Delta_x = 2I_x / (N - 1).\tag{1.19}$$

La fonction $\delta_x(N)$ sert à assurer la convergence du filtre et elle devra respecter les deux conditions suivantes :

$$\begin{aligned}1. \quad & \lim_{N \rightarrow \infty} \delta_x(N) \rightarrow \infty \\ 2. \quad & \lim_{N \rightarrow \infty} \frac{\delta_x(N)}{N} \rightarrow 0.\end{aligned}\tag{1.20}$$

La première condition assure la couverture éventuelle du processus de x_t tandis que la deuxième condition sert à créer un raffinement dans la partition. Il existe une multitude de fonctions $\delta_x(N)$ satisfaisant les deux conditions dont $\delta_x(N) = 5 + \log(N)$. Évidemment, la spécification de cette fonction devra être cohérente avec la distribution de x_t surtout lorsque $\delta p(y_{t-1}, x_{t-1}) \neq 0$. Le but est de couvrir le plus efficacement le domaine de x_t . Cette spécification sera utilisée pour le reste du document et celle-ci n'est pas nécessairement plus valide qu'une autre.

La distribution de la variable latente est continue et devra être découpée en plusieurs intervalles centrés sur chaque $x^{(k)}$. À partir des points \mathbf{X} , il sera possible de

3. Advenant le cas où $\delta p(y_{t-1}, x_{t-1}) = 0$, il est possible de conclure que $\lim_{k \rightarrow \infty} \mathbb{E}[x_{t+k}] = \gamma / (1 - \phi)$ tandis que $\lim_{k \rightarrow \infty} \text{Var}[x_{t+k}] = \sigma_p^2 / (1 - \phi^2)$. Dans le cas contraire, il faudra adapter les équations selon la fonction $\delta p(y_{t-1}, x_{t-1})$.

construire N sous-intervalles $C(\cdot)$ comme étant :

$$\begin{aligned} c(k) &= \frac{x^{(k)} + x^{(k-1)}}{2}, & k = 2, \dots, N \\ c(1) &= -\infty, & c(N+1) = \infty \\ C(k) &= [c(k), c(k+1)], & k = 1, \dots, N, \end{aligned} \quad (1.21)$$

où les bornes $c(1)$ et $c(N+1)$ assurent que l'entièreté de la distribution est couverte et donc, la sommation des probabilités sera toujours égale à un. Ainsi, l'idée de cette construction est que le point $x^{(k)}$ représente toutes les valeurs possibles de l'intervalle $C(k)$. Donc, plus N sera grand, plus l'intervalle sera petit et plus l'approximation sera valide.

Dénotons la matrice $\hat{q}_t^{(i,j)} \forall i, j = 1, \dots, N$ représentant les probabilités du processus x_t de transiter à partir de l'intervalle centré sur $x^{(j)}$ au temps $t-1$ vers l'intervalle $C(i)$ au temps t par

$$\begin{aligned} \hat{q}_t^{(i,j)} &= q(x_t \in C(i) | x_{t-1} \in C(j), y_{t-1}, \theta) \\ &\approx \int_{c(i)}^{c(i+1)} q(x_t | x_{t-1}^{(j)}, y_{t-1}, \theta) dx_t. \end{aligned} \quad (1.22)$$

Cette matrice $N \times N$ sera invariante lorsque $\delta(y_{t-1}, x_{t-1})$ n'est pas fonction de y_{t-1} . Il est possible de considérer un vecteur \mathbf{X} qui tient compte de y_{t-1} pour mieux cibler le domaine de la distribution de x_t mais la construction de l'intervalle devrait pouvoir capter toutes les valeurs lorsque $N \rightarrow \infty$. Par conséquent, la distribution conditionnelle de y_t sachant $x_t \in C(i)$ et $x_{t-1} \in C(j)$ pourra être discrétisée par

$$\begin{aligned} \hat{r}_t^{(i,j)} &\approx r(y_t | x_t \in C(i), x_{t-1} \in C(j), y_{t-1}, \theta) \\ &= r(y_t | x_t^{(i)}, x_{t-1}^{(j)}, y_{t-1}, \theta). \end{aligned} \quad (1.23)$$

1.3.1.3 Algorithme

L'algorithme suppose un vecteur \mathbf{X} fixe et ce dernier s'ajuste facilement dans le cas contraire.

Initialisation ($t = 0$)

- Création du vecteur \mathbf{X} selon l'équation (1.18)
- Calcul du vecteur de probabilités d'observer $x^{(i)} \in C(i)$ à partir de la distribution inconditionnelle comme étant :

$$U_0^i = \int_{c(i)}^{c(i+1)} q(x_0|\theta) dx_0, \quad i = 1, \dots, N.$$

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

- Construction de la matrice $\hat{q}_t^{(i,j)}$ à partir de l'équation (1.22) si $\delta p(y_{t-1}, x_{t-1}) \neq 0$ sinon cette étape se fait dans la partie initialisation parce que la matrice sera invariante.
- Calcul de la matrice ($N \times N$) des probabilités prédictives jointes d'observer $x_t^{(i)} \in C(i)$ et $x_{t-1}^{(j)} \in C(j)$ comme étant :

$$\begin{aligned} P_t^{(i,j)} &= p(x_t^{(i)}, x_{t-1}^{(j)} | y_{t-1}, \theta) \\ &= q(x_t^{(i)} | x_{t-1}^{(j)}, y_{t-1}, \theta) p(x_{t-1}^{(j)} | y_{t-1}, \theta) \\ &\approx \hat{q}_t^{(i,j)} \cdot U_{t-1}^{(j)}. \end{aligned} \tag{1.24}$$

- Calcul de la mise à jour du vecteur de probabilité $x_t^{(i)} \in C(i)$ comme étant :

$$\begin{aligned} U_t^{(i)} &= p(x_t^{(i)} | y_t, \theta) \\ &\approx \frac{\sum_{j=1}^N r(y_t | x_t^{(i)}, x_{t-1}^{(j)}, y_{t-1}, \theta) p(x_t^{(i)}, x_{t-1}^{(j)} | y_{t-1}, \theta)}{p(y_t | y_{t-1}, \theta)} \\ &= \frac{\sum_{j=1}^N \hat{r}_t^{(i,j)} \cdot P_t^{(i,j)}}{f(y_t | y_{t-1}, \theta)}. \end{aligned} \tag{1.25}$$

- Calcul de la vraisemblance d'observer y_t comme étant :

$$\begin{aligned} f(y_{1:T}, \theta) &= f(y_1, \theta) \prod_{t=2}^T f(y_t | y_{1:t-1}, \theta) \\ &\approx f(y_1, \theta) \prod_{t=2}^T f(y_t | y_{t-1}, \theta) \end{aligned} \tag{1.26}$$

où

$$\begin{aligned} f(y_t|y_{t-1}, \theta) &\approx \sum_{i=1}^N \sum_{j=1}^N r(y_t|x_t^{(i)}, x_{t-1}^{(j)}, y_{t-1}, \theta) p(x_t^{(i)}, x_{t-1}^{(j)}|y_{t-1}, \theta) \\ &= \sum_{i=1}^N \sum_{j=1}^N \hat{r}_t^{(i,j)} \cdot P_t^{(i,j)} \end{aligned} \quad (1.27)$$

et

$$f(y_1, \theta) \approx \sum_{i=1}^N \sum_{j=1}^N \hat{r}_1^{(i,j)} \cdot P_1^{(i,j)}.$$

Finalement, l'estimateur θ sera trouvé en résolvant l'équation suivante :

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \log f(y_{1:T}, \theta). \quad (1.28)$$

1.3.2 Filtre à particules (PF)

Les filtres à particules (ou méthodes Monte Carlo séquentielles) sont des procédures ayant comme objectif d'obtenir la densité filtrée $f(x_{0:T}|y_{1:T})$ de la variable latente. Ces méthodes sont principalement basées sur la simulation et convergeront lorsque le nombre de simulations sera infiniment grand. Les propriétés théoriques des PF ont été grandement analysées depuis les travaux de Del Moral (1996). C'est ce dernier qui donna la première preuve robuste du premier filtre à particules de Gordon et al. (1993). Il existe plusieurs excellentes références à ces procédures pour une variété de systèmes à espace-état où la vraisemblance ne peut être écrite de manière analytique. Les références classiques sont Doucet et al. (2001), Arulampalam et al. (2002) et Creal (2012). Récemment, Pitt et al. (2014) ont permis l'estimation de plusieurs modèles à volatilité stochastique avec le filtre à particules dont le modèle SVOL avec des sauts dans l'équation des rendements.

1.3.2.1 Échantillonnage d'importance bayésien (BIS)

L'intuition du filtre à particules est de calculer des quantités du type $\mathbb{E}[h(x_{0:T})|y_{1:T}]$, une espérance d'une fonction des variables d'état conditionnelle aux observations.

Par définition, cette espérance s'écrit :

$$\mathbb{E}[h(x_{0:T})|y_{1:T}] = \int h(x_{0:T})f(x_{0:T}|y_{1:T})dx_{0:T} \quad (1.29)$$

et correspond à une intégrale de $T + 1$ dimensions. En général, il est difficile de simuler à partir de la densité filtrée $f(x_{0:T}|y_{1:T})$ puisqu'elle n'est pas disponible pour tous les systèmes à espace-état. Alors, il faudra utiliser la technique de l'échantillonnage d'importance qui consiste à introduire une densité alternative $f^*(x_{0:T}|y_{1:T})$ avec l'objectif de pouvoir simuler à partir de celle-ci. À l'aide de quelques manipulations, l'équation (1.29) se réécrit comme étant :

$$\begin{aligned} \mathbb{E}[h(x_{0:T})|y_{1:T}] &= \int h(x_{0:T})f(x_{0:T}|y_{1:T})dx_{0:T} \\ &= \int h(x_{0:T})\frac{f(x_{0:T}|y_{1:T})}{f^*(x_{0:T}|y_{1:T})}f^*(x_{0:T}|y_{1:T})dx_{0:T} \\ &= \int h(x_{0:T})\frac{f(y_{1:T}|x_{0:T})f(x_{0:T})}{f^*(x_{0:T}|y_{1:T})f(y_{1:T})}f^*(x_{0:T}|y_{1:T})dx_{0:T} \\ &= \frac{1}{f(y_{1:T})} \int h(x_{0:T})\omega_T(x_{0:T})f^*(x_{0:T}|y_{1:T})dx_{0:T} \end{aligned} \quad (1.30)$$

où le poids non-normalisé ω_T et la fonction de vraisemblance $f(y_{1:T})$ sont :

$$\begin{aligned} \omega_T(x_{0:T}) &= \frac{f(y_{1:T}|x_{0:T})f(x_{0:T})}{f^*(x_{0:T}|y_{1:T})} \\ f(y_{1:T}) &= \int \omega_T(x_{0:T})f^*(x_{0:T}|y_{1:T})dx_{0:T}. \end{aligned} \quad (1.31)$$

En décomposant $f^*(x_{0:T}|y_{1:T})$ par

$$f^*(x_{0:T}|y_{1:T}) = \frac{f^*(y_{1:T}|x_{0:T})f^*(x_{0:T})}{f^*(y_{1:T})} \quad (1.32)$$

permet à l'équation (1.31) d'être calculable. Par conséquent,

$$\begin{aligned} \mathbb{E}[h(x_{0:T})|y_{1:T}] &= \frac{\int h(x_{0:T})\omega_T(x_{0:T})f^*(x_{0:T}|y_{1:T})dx_{0:T}}{\int \omega_T(x_{0:T})f^*(x_{0:T}|y_{1:T})dx_{0:T}} \\ &= \frac{\mathbb{E}^*[h(x_{0:T})\omega_T(x_{0:T})|y_{1:T}]}{\mathbb{E}^*[\omega_T(x_{0:T})|y_{1:T}]} \end{aligned} \quad (1.33)$$

où \mathbb{E}^* signifie l'espérance prise par rapport à f^* et cette espérance sera approximée en simulant N^* réalisations de $x_{0:T}$ appelées particules. L'estimateur Monte Carlo

de cette espérance est calculé comme étant :

$$\hat{\mathbb{E}}[h(x_{0:T})|y_{1:T}] \approx \frac{1}{N^*} \sum_{i=1}^{N^*} h(x_{0:T}^{(i)}) \tilde{\omega}_T^{(i)} \quad (1.34)$$

$$\tilde{\omega}_T^{(i)} = \frac{\omega_T^{(i)}(x_{0:T}^{(i)})}{\frac{1}{N^*} \sum_{i=1}^{N^*} \omega_T^{(i)}(x_{0:T}^{(i)})}$$

Cet algorithme permet de calculer une espérance échantillonnée ajustée en simulant à partir de f^* plutôt que de f . Toutefois, cet algorithme est impraticable en raison de la dimension du problème et la prochaine section présentera une façon de rendre cet algorithme applicable en pratique.

1.3.2.2 Échantillonnage d'importance séquentiel (SIS)

L'échantillonnage d'importance séquentiel sert à réduire la dimension de BIS tout en conservant les fondements de cette procédure. Pour ce faire, trois hypothèses devront être postulées qui sont :

1. les observations courantes et passées de la densité alternative f^* ne dépendent pas des observations futures ;
2. l'évolution temporelle des états est markovienne sachant l'information passée $\mathcal{F}_{t-1} = \{y_1, \dots, y_{t-1}\}$;
3. les observations, sachant les états et l'information passée, sont conditionnellement indépendantes.

Après quelques manipulations, la densité alternative filtrée devient

$$\begin{aligned}
f^*(x_{0:T}|y_{1:T}) &= f^*(x_T, x_{0:T-1}|\mathcal{F}_T) \\
&= f^*(x_T|x_{0:T-1}, \mathcal{F}_T) f^*(x_{0:T-1}|\mathcal{F}_T) \\
&= f^*(x_T|x_{0:T-1}, \mathcal{F}_T) f^*(x_{0:T-1}|\mathcal{F}_{T-1}) \quad (*) \\
&= f^*(x_T|x_{0:T-1}, \mathcal{F}_T) f^*(x_{T-1}|x_{0:T-2}, \mathcal{F}_{T-1}) f^*(x_{0:T-2}|\mathcal{F}_{T-1}) \\
&= \dots \\
&= f(x_0) \prod_{t=1}^T f^*(x_t|x_{0:t-1}, \mathcal{F}_t).
\end{aligned} \tag{1.35}$$

où la ligne (*) est justifiée par l'hypothèse #1. Selon le théorème de Bayes, le numérateur de l'équation (1.31) se réécrit comme

$$f(y_{1:T}|x_{0:T})f(x_{0:T}) = f(y_{1:T}, x_{0:T}) \tag{1.36}$$

et avec quelques manipulation, on obtient

$$\begin{aligned}
f(y_{1:T}, x_{0:T}) &= f(y_{1:T-1}, x_{0:T-1}, x_T, y_T) \\
&= f(x_T, y_T|y_{1:T-1}, x_{0:T-1})f(y_{1:T-1}, x_{0:T-1}) \\
&= f(y_T|x_T, x_{0:T-1}, \mathcal{F}_{T-1})f(x_T|x_{0:T-1}, \mathcal{F}_{T-1})f(y_{1:T-1}, x_{0:T-1}) \\
&= f(y_T|x_{0:T}, \mathcal{F}_{T-1})f(x_T|x_{0:T-1}, \mathcal{F}_{T-1})f(y_{1:T-2}, x_{0:T-2}, x_{T-1}, y_{T-1}) \\
&= \dots \\
&= \prod_{t=1}^T f(y_t|x_{0:t}, \mathcal{F}_{t-1})f(x_t|x_{0:t-1}, \mathcal{F}_{t-1}).
\end{aligned} \tag{1.37}$$

À l'aide des égalités précédentes, cela permet une écriture récursive du poids de l'équation (1.31) comme étant :

$$\begin{aligned}
\omega_T &= \frac{\prod_{t=1}^T f(y_t|x_{0:t}, \mathcal{F}_{t-1})f(x_t|x_{0:t-1}, \mathcal{F}_{t-1})}{\prod_{t=1}^T f^*(x_t|x_{0:t-1}, \mathcal{F}_t)} \\
&= \frac{\prod_{t=1}^{T-1} f(y_t|x_{0:t}, \mathcal{F}_{t-1})f(x_t|x_{0:t-1}, \mathcal{F}_{t-1})}{\prod_{t=1}^{T-1} f^*(x_t|x_{0:t-1}, \mathcal{F}_t)} \left\{ \frac{f(y_T|x_{0:T}, \mathcal{F}_{T-1})f(x_T|x_{0:T-1}, \mathcal{F}_{T-1})}{f^*(x_T|x_{0:T-1}, \mathcal{F}_T)} \right\} \\
&= \omega_{T-1} \frac{f(y_T|x_{0:T}, \mathcal{F}_{T-1})f(x_T|x_{0:T-1}, \mathcal{F}_{T-1})}{f^*(x_T|x_{0:T-1}, \mathcal{F}_T)}.
\end{aligned} \tag{1.38}$$

Cependant, cet algorithme souffre d'un problème de dégénérescence des poids (*weight degeneracy*), car le SIS simplifie l'expression du poids en une équation récursive. Chopin (2004) a démontré que la variance des poids d'importance du SIS croît de manière exponentielle. Après quelques itérations de l'algorithme, une faible minorité de particules auront des poids élevés et le SIS sera ultimement une fonction d'une seule particule.

1.3.2.3 Échantillonnage d'importance séquentiel avec ré-échantillonnage (SISR)

Dans le but de pallier au problème de dégénérescence des poids, une stratégie serait d'appliquer une étape supplémentaire, un ré-échantillonnage (*resampling*) des particules. L'idée consiste à tirer aléatoirement (avec remise) des particules parmi les N^* couples $(x_t^{(i)}, W_t^{(i)})$ où $W_t^{(i)}$ est le poids normalisé d'une certaine particule se dénotant par

$$W_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{k=1}^{N^*} \omega_t^{(k)}}. \quad (1.39)$$

À chaque itération, un nouvel échantillon sera composé en multipliant les particules probables et en éliminant les particules improbables. Cependant, cette stratégie introduit un nouveau problème qui est l'appauvrissement de l'échantillon (*sample impoverishment*). En effet, certaines particules auront des poids élevés et se feront tirer plus fréquemment composant en majorité le nouvel échantillon.

Pour solutionner le problème de l'appauvrissement de l'échantillon, les nouvelles particules devront être tirées aléatoirement au sein d'une distribution continue. Par exemple, un ré-échantillonnage continu basé sur une densité à noyau (*kernel smoothing*) ou le ré-échantillonnage proposé par Pitt (2002) et Malik & Pitt (2011) sont envisageables. Ainsi, à chaque étape de l'algorithme, la distribution de la variable d'état sera représentée par N^* particules différentes ayant des poids normalisés tous significatifs.

1.3.2.4 Choix de la densité alternative filtrée

Selon l'hypothèse #2, la densité f^* est choisie de sorte qu'elle ne dépend que de x_{T-1} et \mathcal{F}_{T-1} ce qui en résulte l'équation du poids suivante :

$$\begin{aligned}\omega_T &= \omega_{T-1} \frac{f(y_T|x_{0:T}, \mathcal{F}_{T-1})f(x_T|x_{T-1}, \mathcal{F}_{T-1})}{f^*(x_T|x_{T-1}, \mathcal{F}_{T-1})} \\ &= \omega_{T-1} f(y_T|x_{0:T}, \mathcal{F}_{T-1}).\end{aligned}\tag{1.40}$$

Ce choix est connu comme étant le « *Bootstrap Particle Filter* » (BPF) et permet de simplifier l'écriture du poids. En lien avec l'hypothèse #3 et des densités de l'équation (1.13), le poids ω_T correspond à

$$\begin{aligned}\omega_T &= \omega_{T-1} f(y_T|x_{0:T}, \mathcal{F}_{T-1}) \\ &= \omega_{T-1} r(y_T|x_T, x_{T-1}, y_{T-1})\end{aligned}\tag{1.41}$$

et le choix de la densité candidate est tout simplement

$$f^*(x_T|x_{T-1}, \mathcal{F}_{T-1}) = q(x_T|x_{T-1}, y_{T-1}).\tag{1.42}$$

Il est également possible d'améliorer la distribution des particules à chaque étape en utilisant l'information venant de la distribution de $f(x_{0:t}^{(i)}|y_{1:t})$. En effet, cela permet d'améliorer la prédiction de la particule x_t conditionnellement à la nouvelle information venant de y_t . L'intuition est d'incorporer les fondements des filtres de Kalman en utilisant une étape supplémentaire de mise à jour. Cette étape supplémentaire ne sera pas utilisée car cette amélioration devient négligeable en utilisant N^* suffisamment grand.

1.3.2.5 Algorithme

L'algorithme décrit est celui du filtre à particules avec ré-échantillonnage (SISR) basé sur une densité à noyau gaussien. À noter qu'il existe d'autres méthodes de ré-échantillonnage plus efficace dans la littérature.

Initialisation (t=0)

- Simulation de $x_0^{(i)} \sim q(x_0|\theta)$, $i = 1, \dots, N^*$ particules à partir de la densité inconditionnelle⁴.

Boucle pour t=1, ..., T

- Simulation de $x_t^{(i)} \sim q(x_t|x_{t-1}^{(i)}, y_{t-1}, \theta)$, $i = 1, \dots, N^*$ particules à partir de la densité candidate.
- Calcul des poids comme étant :

$$\begin{aligned} \omega_t^{(i)} &= \omega_{t-1}^{(i)} r(y_t|x_t^{(i)}, x_{t-1}^{(i)}, y_{t-1}, \theta) \\ &\propto r(y_t|x_t^{(i)}, x_{t-1}^{(i)}, y_{t-1}, \theta) \end{aligned} \quad (1.43)$$

où le signe proportionnel est valable seulement lorsqu'il y a un ré-échantillonnage à chaque étape. Suite au ré-échantillonnage, le poids de toutes les particules devient $\omega_t^{(i)} = 1/N^*$.

- Ré-échantillonnage en deux étapes à partir des N^* couples $(x_t^{(i)}, W_t^{(i)})$ où $W_t^{(i)}$ est le poids normalisé de chaque particule selon l'équation (1.39).

1. Ré-échantillonnage avec remise selon une distribution multinomiale et formation de N^* nouveaux couples où le nouveau poids normalisé sera $1/N^*$ pour toutes les particules.
2. Ré-échantillonnage selon une distribution à noyaux gaussiens à partir des nouveaux couples. Le paramètre de lissage h (*bandwidth*) est

$$h = \left(\frac{4s^5}{3N^*} \right)^{1/5}$$

où s est l'écart-type de l'échantillon des $x_t^{(i)}$ à l'étape (1). Il peut être démontré que cette fenêtre h est celle minimisant le critère de « *mean integrated squared error* » (MISE).

4. L'initialisation de la volatilité pourrait être fixée à une valeur pré-déterminée telle que l'espérance à long terme du processus ou à une valeur observée.

– Posons les nouveaux N^* couples comme étant :

$$\{x_t^{(i)}; \omega_t^{(i)} = W_t^{(i)} = 1/N^*\}$$

où $x_t^{(i)}$ est la nouvelle particule ré-échantillonnée à partir de la distribution à noyaux gaussiens.

Fonction de vraisemblance du SISR

La dernière étape consiste à écrire la fonction de vraisemblance. Hürzeler & Künsch (2001) démontrent qu'un estimateur sans biais de la fonction de vraisemblance est

$$\begin{aligned} \hat{f}(y_{1:T}, \theta) &= \prod_{t=1}^T \bar{\omega}_t \\ \bar{\omega}_t &= \frac{1}{N^*} \sum_{i=1}^{N^*} \omega_t^{(i)} \end{aligned} \tag{1.44}$$

où $\omega_t^{(i)}$ est le poids avant l'étape de ré-échantillonnage. Par la suite, l'estimateur θ sera trouvé en résolvant l'équation suivante :

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \log \hat{f}(y_{1:T}, \theta). \tag{1.45}$$

1.4 Étude Monte Carlo

Les deux méthodes de la section 1.3 servent à approximer numériquement la fonction de vraisemblance. Le but de cette section est de valider (ou d'invalider) que la méthode DNF est une approximation robuste pour la fonction de vraisemblance. La qualité de l'approximation sera analysée de deux façons. Premièrement, la qualité de la fonction de vraisemblance approximée sera vérifiée en plusieurs points aléatoires. La base de comparaison sera le filtre à particules avec un très grand nombre de particules. Deuxièmement, les propriétés statistiques asymptotiques des estimateurs générés par DNF seront étudiées.

1.4.1 Étude de la fonction de vraisemblance

Plusieurs applications sur les filtres à particules démontrent leurs propriétés asymptotiques lorsque $N^* \rightarrow \infty$ et c'est pourquoi il est possible de considérer l'évaluation de la vraisemblance selon la méthode du PF comme étant « exacte ». Ces applications ont été réalisées par plusieurs auteurs dont Del Moral (1996), Doucet et al. (2001), Doucet & Johansen (2008) et Creal (2012). Ainsi, il sera possible de comparer l'approximation du DNF versus PF selon un jeu de paramètres θ et d'une série d'observations $y_{1:T}$. Alors, plusieurs évaluations de la fonction $f(y_{1:T}, \theta)$ permettront de conclure, s'il existe, un potentiel biais sur la fonction.

L'étude se basera sur 50 000 jeux de paramètres tirés aléatoirement sur le domaine plausible de θ . De plus, le modèle étudié est

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \tag{1.46}$$

où $\lambda g(y_{t-1}, x_t, x_{t-1}) = 0$, $\delta p(y_{t-1}, x_{t-1}) = 0$, $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$. Le domaine de θ sera sélectionné pour valider la méthode DNF en économétrie financière. En ce sens, il vaut mieux utiliser une transformation du paramètre γ pour représenter la moyenne à long terme du processus de la variable latente. Alors, en posant u comme étant la moyenne à long terme

$$u = \lim_{k \rightarrow \infty} \mathbb{E}[x_{t+k}] = \gamma / (1 - \phi),$$

l'équation de la variable latente devient

$$x_t = u + \phi(x_{t-1} - u) + \sigma_\nu \nu_t.$$

Selon cette écriture, le paramètre u représente la moyenne du processus de log-variance. Donc, le tirage des 50 000 jeux de paramètres aléatoires se fera sur le

domaine suivant :

$$\begin{aligned}
 \log(0.001^2) &\leq u \leq \log(0.5^2); \\
 0.01 &\leq \phi \leq 0.99; \\
 0.01 &\leq \sigma_\nu \leq 1; \\
 -0.99 &\leq \rho \leq 0.99.
 \end{aligned}
 \tag{1.47}$$

où le domaine de ϕ est nécessaire pour avoir un processus stationnaire tandis que le domaine de σ_ν et ρ est plutôt usuel. Les bornes de u sont choisies de sorte de représenter une pseudo-volatilité d'un sous-jacent incorporant une fréquence journalière, hebdomadaire, mensuelle ou annuelle. Évidemment, il est possible de transformer le paramètre u vers γ pour avoir l'écriture conventionnelle lorsque θ est fixé.

Dénotons par $\theta^{(j)}$, $j = 1, \dots, 50000$, le jeu de paramètres tiré aléatoirement sur le domaine de l'équation (1.47) et par $f(y_{1:T}^{(j)}, \theta^{(j)})$ la trajectoire simulée par ce jeu de paramètre. Il est possible de générer 50 000 trajectoires du processus SVOL chacune ayant 500 observations. Ainsi, il suffit d'évaluer chacune des approximations (DNF et PF) de la fonction de vraisemblance $f(y_{1:500}^{(j)}, \theta^{(j)})$ pour vérifier la présence de biais. Pour négliger les effets liés aux différentes approximations, le nombre de particules et la longueur du vecteur \mathbf{X} seront fixés à $N^* = 25000$ et $N = 150$. Néanmoins, il y aura toujours de légères différences entre les deux méthodes même si ces deux paramètres sont très grands, car ce sont des approximations numériques.

La figure 1.1 représente la correspondance entre la vraisemblance évaluée avec DNF (approximation) et avec PF (quasi-exacte). Ainsi, sur le graphique, il y a une très forte correspondance sur l'ensemble des trajectoires générées et l'histogramme démontre une quasi-égalité numérique en distribution.

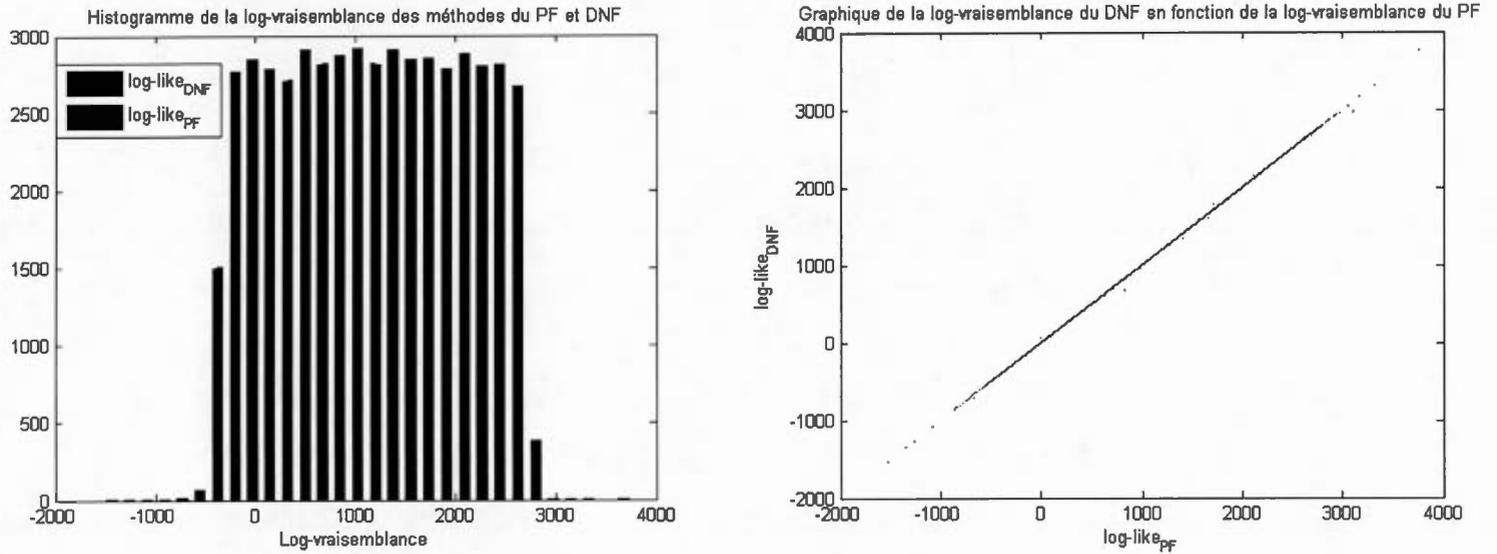


Figure 1.1: Histogramme et graphique des log-vraisemblances des deux méthodes DNF et PF selon 50 000 jeux de paramètres et trajectoires aléatoires.

Cependant, il sera plus judicieux de travailler avec l'erreur relative de la méthode DNF, car elle permet de mieux comparer les différences des évaluations de $\hat{\ell} = \log\{f(y_{1:T}, \theta)\}$ sur une base uni-dimensionnelle. L'erreur relative Δ_{rel} sera dénotée par

$$\Delta_{\text{rel}} = \frac{\hat{\ell}^{(\text{PF})} - \hat{\ell}^{(\text{DNF})}}{\hat{\ell}^{(\text{PF})}} * 100\%. \quad (1.48)$$

La figure 1.2 permet de visualiser les erreurs relatives en fonction de l'approximation de la vraisemblance du DNF. L'évaluation du PF et DNF semblent équivalentes sauf lorsque l'approximation est petite (i.e. $\hat{\ell} \approx 0$). Par exemple, l'erreur relative la plus grande a les résultats suivants :

$$\hat{\ell}^{(\text{PF})} = 0.13353, \quad \hat{\ell}^{(\text{DNF})} = -1.1783, \quad \Delta_{\text{rel}} = 982.4\%.$$

Les résultats démontrent que l'approximation du DNF n'est pas totalement différente de celle du PF, car l'approximation numérique diverge de moins 1.3 points.

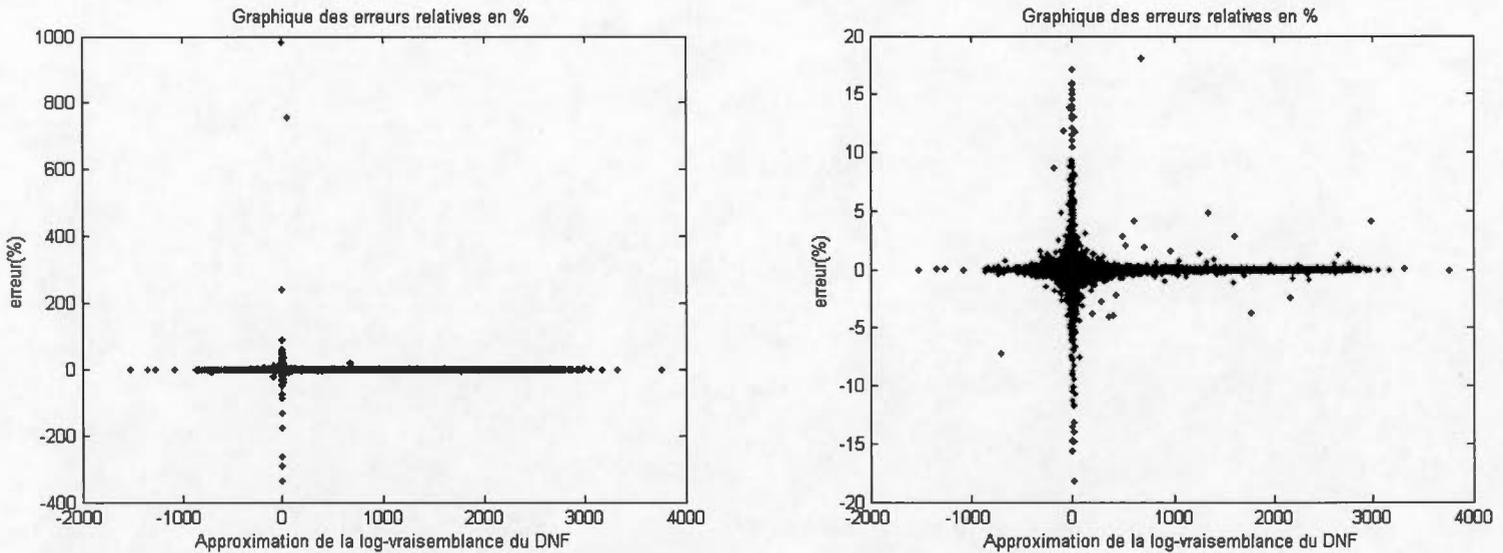


Figure 1.2: Graphique des erreurs relatives avec deux échelles d'axe en « y » différents selon les 50 000 jeux de paramètres et trajectoires aléatoires.

Cependant, sur une base relative, le chiffre indique une évaluation plutôt différente.

D'un autre côté, la grande majorité des erreurs se retrouve dans la zone $\pm 1\%$. De plus, la méthode DNF ne semble pas systématiquement sur-estimer ou sous-estimer la fonction et donc, ne semble pas avoir de biais quelconque. Les erreurs sont uniformément dispersées autour de 0%. Dénotons les erreurs relatives absolues $\Delta_{\text{abs-rel}}$ qui permettront de calculer les quantiles empiriques des erreurs brutes par

$$\Delta_{\text{abs-rel}} = |\Delta_{\text{rel}}|. \quad (1.49)$$

Ainsi, il est possible de visualiser les quantiles empiriques sur le Tableau 1.1. La méthode DNF commet une erreur de plus de 0.11% dans environ 10% des cas et une erreur de plus de 1.4% dans un peu plus de 1% des cas. En général, l'estimation avec DNF semble adéquate et l'approximation est sensiblement la

α	Quantile
0.250	0.003728
0.500	0.011178
0.750	0.035422
0.900	0.113889
0.950	0.245932
0.990	1.408532
0.995	2.964273

Tableau 1.1: Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) selon les 50 000 jeux de paramètres et trajectoires aléatoires.

même que celle du PF. Tel que présenté dans Jacquier et al. (1994), l'ensemble des jeux de paramètres est résumé par le coefficient de variation. La plupart de leurs résultats numériques se basent sur des paramètres ayant des valeurs de ce coefficient égale à 0.1, 1 ou 10. Ils ont ainsi démontré qu'en économétrie financière, le CV du modèle SVOL se situe entre 0.1 et 2. Par conséquent, les 50 000 jeux de paramètres seront classés en fonction de leur CV et la qualité du DNF sera analysée à nouveau, mais en fonction du CV. Ainsi, les mêmes graphiques/analyses seront réalisés en coupant les jeux en quatre groupes notés :

$$\begin{aligned}
 \text{1er groupe : } & \text{CV} \in]0.0 ; 0.1] \\
 \text{2e groupe : } & \text{CV} \in]0.1 ; 1.0] \\
 \text{3e groupe : } & \text{CV} \in]1.0 ; 10] \\
 \text{4e groupe : } & \text{CV} \in]10 ; \infty [.
 \end{aligned}
 \tag{1.50}$$

À partir du Tableau 1.2, il semble que la méthode DNF commet une erreur moins importante lorsque le coefficient variation (CV) est petit. Par conséquent, le DNF est fortement recommandé en économétrie financière par opposition à MM et

Quantiles				
α	1er groupe	2e groupe	3e groupe	4e groupe
0.2500	0.001998	0.003475	0.005518	0.012465
0.5000	0.006817	0.009765	0.015038	0.032272
0.7500	0.022495	0.030074	0.044812	0.093051
0.9000	0.072674	0.097204	0.139360	0.281363
0.9500	0.160983	0.200729	0.308961	0.632189
0.9900	0.829410	1.156442	2.019537	2.938534
0.9950	2.030138	2.304599	4.220605	6.970401
# obs	11714	20889	13934	3463

Tableau 1.2: Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) des quatre groupes en fonction des coefficients de variation selon les intervalles de l'équation (1.50).

QML. Toutefois, l'approximation reste valide même au sein du groupe 4 qui est celui le moins précis. En effet, l'approximation est erronée de plus de 0.28% dans 10% des cas et de plus de 2.93% dans un peu plus de 1% des cas.

À noter que certains CV s'élevaient à plus de $7,8 * 10^{19}$ causant un processus pour la variable latente très volatile. Une hypothèse expliquant ces erreurs parfois élevés serait que le nombre de particules ($N^* = 25000$) du PF ou le nombre de points discrétisant la variable latente ($N = 150$) du DNF ne sont pas suffisamment grands pour obtenir un estimé stable, car les paramètres échantillonnés provoquent un processus très volatil.

1.4.2 Étude Monte Carlo sur les propriétés statistiques du DNF

Dans la section 1.4.1, l'efficacité de la méthode DNF pour approximer la vraisemblance d'une trajectoire donnée a été démontrée numériquement. Maintenant, le but sera de vérifier que le DNF estime bien les paramètres, c'est-à-dire de vérifier que le maximum de vraisemblance a des propriétés statistiques désirables.

Encore une fois, le modèle considéré est

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \tag{1.51}$$

où $\lambda g(y_{t-1}, x_t, x_{t-1}) = 0$, $\delta p(y_{t-1}, x_{t-1}) = 0$, $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$. L'étude se basera sur le jeu de paramètres analysé dans l'article Jacquier et al. (1994) qui est :

$$\gamma = -0.363, \quad \phi = 0.95, \quad \sigma_\nu = 0.26.$$

Dans cet article, ils ne considéraient pas l'effet de levier (i.e. : $\rho = 0$) et ce dernier est généralement négatif sur les marchés financiers. Les mêmes valeurs seront analysées que ceux dans l'article de Clements (2004b) où l'idée était de faire varier celui-ci. En fixant ρ , il sera possible de simuler 1000 trajectoires différentes de 500, 1000 et 2000 observations. Par la suite, il sera possible de calculer le biais moyen et l'écart-type sur chacun des paramètres. Selon les tableaux (1.3 et 1.4), la méthode DNF est approximativement sans biais et converge asymptotiquement vers θ lorsque $T \rightarrow \infty$. De plus, cette méthode n'est pas très sensible aux points initiaux lors de l'optimisation et est plutôt rapide. Il n'est pas nécessaire de fixer $N = 150$ pour avoir une bonne estimation, l'approximation sera valide même lorsque $N = 50$.

	$\gamma = -0.363$	$\phi = 0.95$	$\sigma_\nu = 0.26$	$\rho = 0$
$T = 500$	-0.1268 (0.261)	-0.0174 (0.034)	0.0087 (0.067)	-0.0035 (0.153)
$T = 1000$	-0.0567 (0.146)	-0.0078 (0.020)	0.0027 (0.044)	-0.0033 (0.101)
$T = 2000$	-0.0259 (0.095)	-0.0036 (0.013)	-0.0006 (0.031)	0.0007 (0.076)
	$\gamma = -0.363$	$\phi = 0.95$	$\sigma_\nu = 0.26$	$\rho = -0.3$
$T = 500$	-0.1147 (0.301)	-0.0158 (0.042)	0.0052 (0.061)	-0.0110 (0.149)
$T = 1000$	-0.0458 (0.139)	-0.0063 (0.0190)	0.0012 (0.044)	-0.0018 (0.097)
$T = 2000$	-0.0252 (0.090)	-0.0034 (0.012)	-0.0007 (0.030)	0.0052 (0.072)

Tableau 1.3: Résultats numériques de l'étude de la convergence du DNF (1ère partie : $\rho = 0$ et $\rho = -0.3$) où le biais moyen est répertorié ainsi que l'écart-type entre parenthèse.

	$\gamma = -0.363$	$\phi = 0.95$	$\sigma_\nu = 0.26$	$\rho = -0.5$
$T = 500$	-0.0876 (0.222)	-0.0121 (0.031)	0.0009 (0.055)	-0.0180 (0.137)
$T = 1000$	-0.0384 (0.116)	-0.0053 (0.016)	-0.0009 (0.036)	-0.0081 (0.089)
$T = 2000$	-0.0185 (0.079)	-0.0025 (0.011)	-0.0022 (0.027)	0.0010 (0.064)
	$\gamma = -0.363$	$\phi = 0.95$	$\sigma_\nu = 0.26$	$\rho = -0.7$
$T = 500$	-0.0516 (0.158)	-0.0072 (0.022)	-0.0044 (0.047)	-0.0242 (0.109)
$T = 1000$	-0.0212 (0.092)	-0.0029 (0.013)	-0.0032 (0.032)	-0.0081 (0.071)
$T = 2000$	-0.0115 (0.064)	-0.0015 (0.009)	-0.0029 (0.023)	0.0014 (0.048)

Tableau 1.4: Résultats numériques de l'étude de la convergence du DNF (2e partie : $\rho = -0.5$ et $\rho = -0.7$) où le biais moyen est répertorié ainsi que l'écart-type entre parenthèse.

1.4.3 Conclusion

Les filtres à particules sont des méthodes exigeantes en temps de calcul ayant des problèmes numériques de convergence lors de la maximisation de $f(y_{1:T}, \theta)$. Les problèmes de convergence sont causés par les nombreux maximums locaux qui se créent dans la fonction de vraisemblance. En effet, la fonction de vraisemblance est irrégulière, car cette dernière est obtenue par simulation. L'utilisation du ré-échantillonnage continue crée tout de même des sauts et ils ne peuvent être éliminés. Ainsi, les points initiaux vont grandement affecter le résultat de l'optimisateur et ce même si le germe des simulations est fixé lors de cette étape. Il est nécessaire d'avoir des méthodes de recherche d'optimum globaux robuste et d'avoir un nombre de particules très grands (200 000 et plus). Toutefois, les filtres à particules peuvent être utilisés pour obtenir une bonne approximation de la fonction de vraisemblance.

À la lumière de ces résultats, la méthode DNF effectue une approximation suffisamment précise dans un délai raisonnable. C'est pourquoi l'utilisation de cette méthode est une solution intéressante puisqu'elle est également rapide et simple.

Dans l'éventualité où le temps de calcul n'est pas une contrainte, une idée serait de cibler le domaine de θ avec la méthode DNF. Ensuite, cette information serait judicieusement utilisée pour permettre une estimation plus efficace via le filtre à particules sachant qu'il est très délicat d'utiliser cette technique dans le contexte de maximum du vraisemblance.

1.5 Exemple empirique

Dans cette section, la qualité de l'adéquation de différents modèles SVOL sera analysée ainsi que d'autres modèles couramment utilisés dans l'évaluation de fonds

distincts. Les paramètres des modèles seront estimés à partir des bases de données de l'indice composé du S&P500 et l'indice composé du S&P/TSX⁵. Les deux séries ont 1825 (respectivement 1826) observations hebdomadaires ($\Delta_t = 1/52$) datant du 1er janvier 1980 au 1er janvier 2015. L'écriture des modèles sera annualisée en introduisant la fréquence d'observation Δ_t pour assurer une uniformité à travers les résultats.

Les critères d'information de Aikake (AIC) et bayésien (BIC) permettront de classer les modèles entre eux et se notent :

$$\begin{aligned} \text{AIC} &= -2 \log f(y_{1:T}, \theta) + 2(p) \\ \text{BIC} &= -2 \log f(y_{1:T}, \theta) + p \log(n) \end{aligned} \quad (1.52)$$

où p est le nombre de paramètres et n le nombre d'observations. Il est important de noter qu'il s'agit d'une analyse simplifiée de la qualité relative des modèles. Une analyse plus exhaustive comprendrait des tests de normalité (Ljung-Box, Jarque-Bera, Shapiro-Wilk, ARCH-LM, etc.) sur les résidus engendrés par les modèles et la reproduction de faits stylisés observés sur les marchés (changement de régimes, queue de gauche lourde et asymétrique, volatilité stochastique, sauts, retour à la moyenne, etc).

1.5.1 Comparaison entre des variantes du SVOL

Le modèle SVOL analysé aura les spécifications suivantes :

$$\begin{aligned} y_t &= \left(\lambda - \frac{1}{2}\sigma_t^2\right)\Delta_t + \sqrt{\Delta_t}\sigma_t\epsilon_t \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned} \quad (1.53)$$

où $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda - \frac{1}{2}\sigma_t^2$, $\delta p(y_{t-1}, x_{t-1}) = 0$, $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$. Pour visualiser l'apport de l'effet de levier, il y aura un cas où $\rho = 0$ et un autre où $\rho \neq 0$.

5. Les deux bases de données (S&P500 et S&P/TSX) sont obtenues à partir de Yahoo! Finance.

S&P/TSX	SVOL ($\rho = 0$)	SVOL ($\rho \neq 0$)	SVOL (AAA)
γ/τ	-0.18139	-0.17786	0.13120
ϕ	0.95526	0.95612	0.04719
σ_v	0.24387	0.24653	0.12559
ρ	0	-0.54364	-0.56977
λ/A	0.12519	0.12607	0.23561
B	x	x	-1.87103
C	x	x	6.53784
$\log f(y_{1:T}, \theta)$	4618.76	4649.04	4650.06
AIC	-9229.5	-9288.1	-9286.1
BIC	-9207.5	-9260.5	-9247.6

Tableau 1.5: Résultats du maximum de vraisemblance pour les trois modèles SVOL ayant des versions différentes selon l'indice composé du S&P/TSX.

De plus, la version du modèle AAA sera également estimée. Cette variante est susceptible d'avoir une meilleure adéquation, car la fonction $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda_1 + \lambda_2 \sigma_t + \lambda_3 \sigma_t^2$ est un cas particulier de la dérive spécifiée du SVOL lorsque $\lambda_2 = (\lambda_3 + 0.5) = 0$. Les résultats du maximum de vraisemblance obtenus avec le DNF peuvent être visualisés dans le Tableau 1.5 et 1.6 . Le modèle SVOL avec corrélation est celui minimisant tous les critères d'information pour les deux séries de données sauf pour un cas (i.e : AIC - S&P/TSX) et donc, il a la meilleure adéquation parmi les modèles investigués. Le coefficient associé à la présence de l'effet de levier est très élevé pour les deux versions du modèle SVOL argumentant l'importance de ce fait stylisé. Ce résultat est donc cohérent avec les articles de Black (1976), Christie (1982) et Tsay (2001).

S&P500	SVOL ($\rho = 0$)	SVOL ($\rho \neq 0$)	SVOL (AAA)
γ / τ	-0.15851	-0.15952	0.13915
ϕ	0.95991	0.95921	0.05560
σ_v	0.21828	0.24200	0.12590
ρ	0	-0.63761	-0.71013
λ/A	0.14308	0.14529	-0.09517
B	x	x	2.24598
C	x	x	-2.21275
$\log f(y_{1:T}, \theta)$	-4522.64	4558.66	4562.43
AIC	-9037.2	-9107.3	-9110.9
BIC	-9015.2	-9079.8	-9072.3

Tableau 1.6: Résultats du maximum de vraisemblance pour les trois modèles SVOL ayant des versions différentes selon l'indice composé du S&P500.

1.5.2 Comparaison avec des modèles financiers populaires

Dans un deuxième temps, la comparaison se fera avec d'autres modèles connus dans la littérature. Il y aura cinq modèles sélectionnés en lien avec les articles de Hardy et al. (2006) et du rapport du Canadian Institute of Actuaries (2012) qui les ont également analysés dans le contexte des fonds distincts. Les cinq autres modèles sont :

1. le processus log-normal (BS) introduit par Black & Scholes (1973) défini par

$$y_t = \left(\mu_1 - \frac{1}{2}\sigma_1^2\right)\Delta_t + \sqrt{\Delta_t}\sigma_1\epsilon_t \quad (1.54)$$

où ϵ_t est un bruit gaussien ;

2. le processus GARCH(1,1) introduit par Bollerslev (1986) défini par

$$y_t = \left(\mu_1 - \frac{1}{2}\sigma_t^2\right)\Delta_t + \sqrt{\Delta_t}\sigma_t\epsilon_t \quad (1.55)$$

$$\sigma_t^2 = \gamma_1 + \alpha_1(\sigma_{t-1}\epsilon_{t-1})^2 + \beta_1\sigma_{t-1}^2$$

où $\gamma_1 \geq 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$ et ϵ_t est un bruit gaussien ;

3. le processus log-normal à changement de régimes RSLN (2 états) introduit par Hamilton (1989) défini par

$$y_t = \left(\mu_{R_t} - \frac{1}{2}\sigma_{R_t}^2\right)\Delta_t + \sqrt{\Delta_t}\sigma_{R_t}\epsilon_t \quad (1.56)$$

où le processus latent $R_t = 1, 2$ (le régime) est une chaîne de Markov invariante s'écrivant

$$\begin{pmatrix} p_{11}^{\Delta_t} & p_{12} = 1 - p_{11}^{\Delta_t} \\ p_{21} = 1 - p_{22}^{\Delta_t} & p_{22}^{\Delta_t} \end{pmatrix} \quad (1.57)$$

où p_{ij} signifie la probabilité de transiter du régime i vers j ;

4. le processus Merton Jump-Diffusion (MJD) introduit par Merton (1976) défini par

$$y_t = \left(\mu_1 - \beta\alpha - 0.5\sigma_1^2\right)\Delta_t + \sqrt{\Delta_t}\sigma_1\epsilon_t + \sum_{i=1}^{N_t} \log(1 + X^{(i)}) \quad (1.58)$$

$$X^{(i)} = \log(1 + \beta) - 0.5\sigma_2^2 + \sigma_2\nu_t^{(i)}$$

où N_t est un processus de Poisson d'intensité $\alpha\Delta_t$ et ν_t et ϵ_t sont des bruits gaussiens indépendants ;

5. le processus GARCH à changement de régimes MS-GARCH (2 états) introduit par Hamilton and Susmel (1994)⁶ défini par

$$\begin{aligned} y_t &= \mu_{R_t} + \sigma_t(R_{1:t})\epsilon_t \\ \sigma_t^2(R_{1:t}) &= \omega_{R_t} + \alpha_{R_t}\eta_{t-1}^2(R_{t-1}) + \beta_{R_t}\sigma_{t-1}^2(R_{1:t-1}) \\ \eta_{t-1}(R_{t-1}) &= y_{t-1} - \mu_{R_{t-1}} \end{aligned} \quad (1.59)$$

où $\mu_{R_t} = \mathbb{E}[y_t|R_t]$ et $\sigma_t^2(R_{1:t}) = \text{Var}[y_t|y_{1:t-1}, R_{1:t}]$ sont respectivement la moyenne conditionnelle de y_t et la variance conditionnelle sachant tout l'évolution du processus latent $R_{1:t}$. Le processus latent R_t (le régime) est une chaîne de Markov invariante s'écrivant

$$\begin{pmatrix} p_{11}^{\Delta t} & p_{12} = 1 - p_{11}^{\Delta t} \\ p_{21} = 1 - p_{22}^{\Delta t} & p_{22}^{\Delta t} \end{pmatrix} \quad (1.60)$$

avec p_{ij} signifiant la probabilité de transiter du régime i vers j . Pour simplifier le problème d'estimation du MS-GARCH, il est souvent argumenté de poser $\alpha_{R_t} = \alpha$ et $\beta_{R_t} = \beta$ pour $R_t = 1, 2$, mais les paramètres μ_{R_t} et ω_{R_t} sont définis par le régime au temps t .

Les résultats du maximum de vraisemblance des trois modèles sont répertoriés dans le tableau 1.7. Le modèle SVOL est celui minimisant les critères d'information. En ce sens, la présence de l'effet de levier et une adéquation d'une volatilité latente sembleraient être des faits stylisés importants dans le cadre de

6. Plusieurs techniques d'estimation ont été avancées dans la littérature pour ce modèle dont la première par Gray (1996) pour résoudre le problème de « *path dependence* ». Dans ce mémoire, la technique utilisée permettant une estimation efficace du MS-GARCH sera celle décrite dans Augustyniak, Boudreault et Morales (2015).

S&P/TSX	BS	GARCH(1,1)	RSLN(2 états)	MJD	MS-GARCH (2 états)
$\log f(y_{1:T}, \theta)$	4342.66	4585.21	4596.22	4512.85	4633.57
AIC	-8681.3	-9162.4	-9180.4	-9015.7	-9251.1
BIC	-8670.3	9140.3	-9147.4	-8988.2	-9207.1
S&P500	BS	GARCH(1,1)	RSLN(2 états)	MJD	MS-GARCH (2 états)
$\log f(y_{1:T}, \theta)$	4300.21	4498.11	4489.88	4415.42	4515.52
AIC	-8596.4	-8988.2	-8967.8	-8820.9	-9015.0
BIC	-8585.3	-8966.1	-8934.7	-8793.3	-8971.0

Tableau 1.7: Résultats du maximum de vraisemblance pour les cinq modèles stochastiques selon les indices composés du S&P/TSX et S&P500. Pour les détails sur les paramètres, se référer au tableau A.1 dans l'annexe.

la modélisation des rendements boursiers. Ces conclusions ont été réalisées avec $\delta p(y_{t-1}, x_{t-1}) = 0$ et $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda - 0.5\sigma_t^2$ qui sont les spécifications les plus simples pour l'adéquation de la log-variance et la dérive des rendements.

CHAPITRE II

ESTIMATION DU MODÈLE À VOLATILITÉ STOCHASTIQUE EN TEMPS CONTINU (HESTON)

2.1 Le modèle à volatilité stochastique (temps continu)

Le processus à volatilité stochastique avec une variable latente x_t à temps continu (ou le modèle de Heston (1993)) est

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dW_t^{(1)} \\ d\sigma_t^2 &= dx_t = \kappa(\theta - x_t)dt + \sigma\sqrt{x_t}dW_t^{(2)} \end{aligned} \quad (2.1)$$

où κ est la vitesse de retour à la moyenne du processus de variance, θ est la moyenne à long terme du processus de variance, σ est la volatilité du processus de variance, μ est la dérive stationnaire du sous-jacent S_t , $W_t^{(1)}$ et $W_t^{(2)}$ sont des mouvements browniens standards.

Cette écriture fait référence à des équations différentielles stochastiques (EDS) et elle permet de définir le comportement du processus sur un intervalle de longueur infinitésimale $[t, t + dt]$. Il est supposé que le prix du sous-jacent est observé sur une base continue ce qui n'est pas le cas du modèle SVOL.

Tout comme le modèle SVOL, le modèle de Heston peut incorporer l'effet de levier en supposant que les mouvements browniens standard $W_t^{(1)}$ et $W_t^{(2)}$ sont mutuellement corrélés de sorte que $\mathbb{E}[W_t^{(1)} \cdot W_t^{(2)}] = \rho$.

La principale différence entre le modèle de Heston et le modèle SVOL est l'adéquation de la composante latente. En effet, le modèle de Heston représente la variance comme étant un processus de Cox Ingersoll Ross (CIR) tandis que le modèle SVOL modélise la log-variance comme étant un processus de Ornstein–Uhlenbeck (OU) avec une discrétisation d'Euler. Dans les deux cas, la variance sera toujours positive. En effet, la variance du SVOL est obtenue en prenant l'exponentiel de x_t , une fonction positive. Cependant, dans le modèle de Heston, Feller (1951) a démontré que le processus de CIR sera strictement positif lorsque la condition suivante est

respectée

$$2\kappa\theta > \sigma^2. \quad (2.2)$$

Dans la littérature, l'équation (2.2) est référée comme étant la condition de Feller et évite le processus de variance de s'affaisser à « 0 ».

Pour estimer les paramètres, il vaut mieux travailler avec les log-rendements via la transformation $y_t = \log S_t$. Avec le lemme d'Itô, le modèle de Heston devient

$$\begin{aligned} dy_t &= \left(\mu - \frac{1}{2}x_t \right) dt + \sqrt{x_t}dW_t^{(1)} \\ dx_t &= \kappa(\theta - x_t)dt + \sigma\sqrt{x_t}dW_t^{(2)} \end{aligned} \quad (2.3)$$

où $\Omega = \{\kappa, \theta, \sigma, \rho, \mu\}$ est un vecteur de paramètres.

2.1.1 Le modèle de Heston sous la mesure de probabilité \mathbb{Q}

La mesure de probabilité neutre au risque \mathbb{Q} sert ultimement à tarifier des produits dérivés. Dans la littérature, il existe un passage du modèle sous la mesure de probabilités physique \mathbb{P} vers la mesure neutre au risque. L'objectif est d'évaluer des produits dérivés sur le prix d'un actif risqué S_t en absence d'arbitrage. Le théorème fondamental d'évaluation stipule qu'en absence d'arbitrage il existe au moins une mesure neutre au risque \mathbb{Q} . Le modèle est défini dans un marché incomplet car la variance évolue de façon stochastique dans le temps et cette dernière n'est pas transigée. Par conséquent, un marché incomplet implique généralement une infinité de prix possibles pour le même produit dérivé ce qui n'est pas pratique. Alors, la solution de Heston est de proposer un changement de mesure unique dans une certaine classe pour chacun des deux mouvements browniens. Le premier changement est

$$\tilde{W}_t^{(1)} = W_t^{(1)} + \int_0^t \frac{\mu - r}{\sqrt{x_s}} ds. \quad (2.4)$$

où r est le taux sans risque et le deuxième est

$$\tilde{W}_t^{(2)} = W_t^{(2)} + \int_0^t \frac{\lambda^* x_s}{\sigma \sqrt{x_s}} ds. \quad (2.5)$$

En invoquant le théorème de Girsanov (et la condition de Novikov), on peut démontrer que sous \mathbb{Q} , le mouvement brownien transformé est toujours un mouvement brownien standard. Par conséquent, il est pratique de poser

$$\kappa^* = \kappa + \lambda^*, \quad \theta^* = \frac{\kappa\theta}{\kappa + \lambda^*}, \quad (2.6)$$

pour déduire le modèle sous la mesure neutre au risque et il s'écrit

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{x_t} S_t d\tilde{W}_t^{(1)} \\ dx_t &= \kappa^*(\theta^* - x_t) dt + \sigma \sqrt{x_t} d\tilde{W}_t^{(2)}. \end{aligned} \quad (2.7)$$

De plus, les deux mouvements browniens standards sous \mathbb{Q} sont toujours mutuellement corrélés (i.e. : $\mathbb{E}[\tilde{W}_t^{(1)} \cdot \tilde{W}_t^{(2)}] = \rho$).

2.2 Synthèse des techniques d'estimation

Il existe deux classes de techniques statistiques pour obtenir des estimateurs de Ω dans le modèle de Heston : la calibration et l'estimation. La calibration est plus populaire en pratique tandis que l'estimation au sens statistique du terme (MM, MLE, Bayésien, etc.) est davantage utilisée dans la littérature économétrique. La distinction entre les deux techniques est liée par l'utilisation de différentes observations pour déduire les paramètres. Généralement, les deux méthodes utilisent les prix d'un sous-jacent, mais la calibration nécessite également les prix de produits dérivés du sous-jacent. Les données utilisées pour inférer les paramètres vont avoir un impact, car elles ne sont pas conceptuellement sous la même mesure de probabilités.

Il a été démontré, dans les sections 2.1 et 2.1.1 que le modèle peut s'écrire sous

les deux mesures de probabilités équivalentes \mathbb{P} et \mathbb{Q} . Cela permet d'effectuer un passage d'une mesure de probabilité vers une autre lorsqu'une hypothèse est prise quant à la prime de risque $\lambda(S_t, x_t, t)$. Donc, il est toujours possible d'obtenir les paramètres du modèle sous \mathbb{P} à partir des paramètres sous \mathbb{Q} et vice versa. Cependant, le contexte de l'utilisation du modèle devrait être sous la même mesure de probabilité implicite à la méthode statistique utilisée, car cela permet de rester cohérent. La calibration devrait être utilisée pour tarifier des produits dérivés tandis que l'estimation devrait être utilisée pour projeter un actif risqué.

Plus récemment, les techniques d'estimation et de calibration se confondent davantage. En effet, il y a des techniques basées sur le maximum de vraisemblance pour obtenir la dynamique du modèle sous la mesure \mathbb{P} tout en utilisant simultanément les prix de produits dérivés pour déduire la prime de risque avec la version sous \mathbb{Q} (voir Christofferson (2010)).

2.2.1 Calibration

La tarification d'options doit se faire sous la mesure neutre au risque \mathbb{Q} . L'idée de la calibration consiste à minimiser une distance entre des quantités observées et des quantités théoriques d'un modèle paramétrique via une fonction de pénalité quelconque. Très souvent, deux types de quantités sont utilisées pour effectuer cette minimisation soit les prix d'options ou les volatilités théoriques de ceux-ci. Ainsi, les paramètres obtenus seront ceux sous la mesure \mathbb{Q} , car cette technique infère ceux-ci à partir des prix d'options qui sont supposés tarifés sous cette mesure.

Dans le modèle de Heston, il existe des formules semi-fermées pour calculer les prix d'options vanilles (voir Heston 1993) et les prix peuvent être observés sur les marchés financiers selon différentes caractéristiques (échéance, « moneyness », etc.). Par conséquent, les techniques de calibration sont envisageables d'un point

vue computationnel parce qu'il n'est pas nécessaire d'avoir recours à la simulation Monte Carlo.

Dénotons $i = 1, \dots, n$ options vanilles chacune ayant les caractéristiques $\{\tau_i, K_i, C_i(S_t, t)^{(\text{obs})}\}$ où τ_i , K_i et $C_i(S_t, t)^{(\text{obs})}$ représente respectivement l'échéance, le prix d'exercice et le prix de l'option vanille. En fixant le jeu de paramètres Ω du modèle de Heston, le prix théorique correspondant aux caractéristiques de l'option i sera dénoté par $C_i(S_t, t)^{(\text{theo-Hes})}$.

2.2.1.1 Fonction de pénalité sur les prix des produits dérivés

La première catégorie consiste à minimiser une distance entre les prix de produits dérivés théoriques au modèle et ceux observés via une fonction de pénalité g . Par exemple, minimiser la moyenne des erreurs quadratiques (MSE) consiste à trouver les paramètres Ω résolvant le système suivant :

$$\underset{\Omega \in \Omega}{\operatorname{argmin}} \quad \frac{1}{n} \sum_{i=1}^n \omega_i \times g \left\{ C_i(S_t, t)^{(\text{obs})}, C_i(S_t, t)^{(\text{theo-Hes})} \right\} \quad (2.8)$$

où

$$g \left\{ C_i(S_t, t)^{(\text{obs})}, C_i(S_t, t)^{(\text{theo-Hes})} \right\} = \left[C_i(S_t, t)^{(\text{obs})} - C_i(S_t, t)^{(\text{theo-Hes})} \right]^2 \quad (2.9)$$

et ω_i est un poids optionnel sur chaque option. Le système peut être aussi résolu sous une base relative (l'erreur moyenne des erreurs quadratiques relatives) et consiste à trouver

$$\underset{\Omega \in \Omega}{\operatorname{argmin}} \quad \frac{1}{n} \sum_{i=1}^n \omega_i \frac{g \left\{ C_i(S_t, t)^{(\text{obs})}, C_i(S_t, t)^{(\text{theo-Hes})} \right\}}{C_i(S_t, t)^{(\text{obs})}}. \quad (2.10)$$

où g est défini par l'équation (2.9). Un autre choix serait de travailler avec l'erreur absolue en posant

$$g \left\{ C_i(S_t, t)^{(\text{obs})}, C_i(S_t, t)^{(\text{theo-Hes})} \right\} = \left| C_i(S_t, t)^{(\text{obs})} - C_i(S_t, t)^{(\text{theo-Hes})} \right| \quad (2.11)$$

au lieu de l'erreur quadratique.

Un problème survient en travaillant avec les fonctions de pénalité sur les prix d'option. La grandeur des prix de produits dérivés dépend des caractéristiques sélectionnées (échéance, prix d'exercice, etc.). Dans le format de l'équation (2.8), les options ayant des prix élevés vont avoir plus d'impact pour un même pourcentage d'erreur que ceux ayant des prix faibles. Très souvent, les options de longue échéance ont des prix élevés. Les paramètres résultant de cette minimisation vont possiblement refléter une dynamique à long terme.

D'un autre côté, le format de l'équation (2.10) sera influencée par les prix d'option de courte échéance à cause du dénominateur. Encore une fois, les paramètres pourraient être influencés et représenter davantage une dynamique à court terme. Cela souligne l'importance d'assigner un poids ω_i de sorte que toutes les options aient un impact similaire lors de la minimisation.

2.2.1.2 Fonction de pénalité sur les volatilités théoriques

La deuxième catégorie consiste à minimiser les erreurs entre les volatilités théoriques observées sur les marchés $IV^{(\text{obs})}$ et celles théoriques au modèle de Heston $IV^{(\text{theo-Hes})}$ via la fonction de pénalité g . Encore une fois, il est possible de minimiser une quelconque distance (erreur absolue, quadratique, relative) entre ces deux quantités. Par exemple, minimiser la moyenne des erreurs quadratiques des volatilités théoriques est

$$\operatorname{argmin}_{\Omega \in \Omega} \frac{1}{n} \sum_{i=1}^n \omega_i \left[IV_i^{(\text{obs})} - IV_i^{(\text{theo-Hes})} \right]^2. \quad (2.12)$$

Il est recommandé d'utiliser ces quantités, car elles sont très souvent de même ordre peu importe les caractéristiques des options. Cependant, le désavantage est relié à la complexité supplémentaire du problème. En effet, il faut trouver les prix

théoriques du modèle de Heston $C_i(S_t, t)^{(\text{theo-Hes})}$ et ensuite trouver les volatilités théoriques $IV_i^{(\text{theo-Hes})}$ du modèle de Black-Scholes à partir de $C_i(S_t, t)^{(\text{theo-Hes})}$ avec un algorithme quelconque de recherche de zéro.

L'expansion de Lewis (2000) peut être utilisé pour approximer la volatilité théorique du modèle de Heston et éviter l'étape de bisection. Un autre moyen de rendre les calculs plus efficaces est d'utiliser l'approximation décrite dans Christoffersen & al. (2009). Cela permet d'approximer l'équation (2.12) via l'équation (2.8) en posant ω_i égale à la dérivée « *vega* » au carré du modèle de Black-Scholes.

Bakshi & al. (1997), Bams & al. (2009), Christoffersen & Jacobs (2004) et plusieurs autres auteurs ont abordé la problématique de l'estimation du modèle de Heston avec une technique de calibration. Les conclusions ne semblent pas indiquer un problème de minimisation plus valable qu'un autre.

2.2.2 Estimation statistique

L'estimation repose sur des méthodes statistiques où les paramètres Ω sont obtenus seulement à partir de prix observés du sous-jacent. Tel que montré dans le chapitre 1, il existe plusieurs méthodes (MM, MCMC, etc.) mais la plus populaire consiste à maximiser la fonction de vraisemblance (ML). Dans le cas du modèle de Heston, il est impossible d'écrire cette fonction pour les mêmes raisons que pour le modèle SVOL. Il faudra utiliser des méthodes numériques pour résoudre l'intégrale de dimension $T + 1$.

2.2.2.1 Atiya & Wall (2009)

Pour y arriver, les auteurs décomposent la fonction de vraisemblance de la variance au temps T à partir d'une équation récursive. Par la suite, l'idée est de déduire la

fonction de vraisemblance d'observer $y_{1:T}$ à partir de la fonction de vraisemblance de la variable latente x_T . Ceci est représenté par

$$\begin{aligned}\mathcal{L}_T(x_T) &= p(x_T|y_{1:T}) \\ &= \frac{p(x_T, y_{1:T})}{p(y_{1:T})}\end{aligned}\quad (2.13)$$

et donc, la vraisemblance sera obtenue avec l'égalité suivante

$$p(y_{1:T}) = \int p(x_T, y_{1:T}) dx_T \quad (2.14)$$

suite à une approximation de $p(x_T, y_{1:T})$ à l'aide d'une formule récursive.

Premièrement, à l'aide de l'équation (2.3), la distribution jointe de y_{t+1} et x_{t+1} sachant y_t et x_t est

$$p(y_{t+1}, x_{t+1}|y_t, x_t) \approx \Phi(\mu_{t+1}, \Sigma_{t+1}) \quad (2.15)$$

où Φ est la densité d'une normale bivariée, Δ_t est l'intervalle entre deux observations et

$$\mu_{t+1} = \begin{pmatrix} y_t + (\mu - \frac{1}{2}x_t)\Delta_t \\ x_t + \kappa(\theta - x_t)\Delta_t \end{pmatrix}, \quad \Sigma_{t+1} = x_t \Delta_t \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \quad (2.16)$$

sont respectivement les matrices de la moyenne et variance-covariance de cette densité.

Deuxièmement, en reprenant l'équation (2.13) au temps $t+1$ et à une constante près qui ne dépend pas de la variable d'intérêt x_t , il est possible de déduire que

$$\begin{aligned}\mathcal{L}_{t+1}(x_{t+1}) &\propto p(x_{t+1}, y_{1:t+1}) \\ &= \int p(x_{t+1}, y_{1:t+1}, x_{0:t}) dx_{0:t} \\ &= \int p(x_{t+1}, y_{t+1}|x_{0:t}, y_{1:t}) p(x_{0:t}, y_{1:t}) dx_{0:t-1} dx_t \\ &\propto \int p(y_{t+1}, x_{t+1}|y_t, x_t) \mathcal{L}_t(x_t) dx_t, \quad (*)\end{aligned}\quad (2.17)$$

$$\mathcal{L}_{t+1}(x_{t+1}) \propto d_t (ab_t)^{-1/4} e^{-2\sqrt{ab_t}} \mathcal{L}_t \left(\sqrt{\frac{b_t}{a}} \right)$$

où (*) est argumenté par la propriété markovienne du modèle et

$$\begin{aligned}
 a &= \frac{\kappa' \kappa' + \rho \sigma \kappa' \Delta_t + 0.25 \sigma^2 \Delta_t^2}{2\sigma^2(1 - \rho^2)\Delta_t} \\
 b_t &= \frac{(x_{t+1} - \alpha \Delta_t)^2 - 2\rho\sigma(x_{t+1} - \alpha \Delta_t)(\Delta y_{t+1} - \mu \Delta_t) + \sigma^2(\Delta y_{t+1} - \mu \Delta_t)^2}{2\sigma^2(1 - \rho^2)\Delta_t} \\
 d_t &= \frac{1}{2\pi\sigma\sqrt{1 - \rho^2}\Delta_t} \exp\left(\frac{(2\kappa' + \rho\sigma\Delta_t)(x_{t+1} - \alpha \Delta_t) - (2\rho\sigma\kappa' + \sigma^2\Delta_t)(\Delta y_{t+1} - \mu \Delta_t)}{2\sigma^2(1 - \rho^2)\Delta_t}\right) \\
 \Delta y_{t+1} &= y_{t+1} - y_t, \quad \kappa' = 1 - \kappa \Delta_t, \quad \alpha = \kappa \theta.
 \end{aligned} \tag{2.18}$$

L'équation (2.17) dépend de la valeur x_{t+1} et cette valeur est déduite à partir de l'égalité

$$x_{t+1} = \sqrt{\frac{b_{t+1}}{a}}. \tag{2.19}$$

Cette égalité est solutionnée en trouvant la racine quadratique du système suivant :

$$\begin{aligned}
 x_{t+1} &= \sqrt{B^2 - C} - B \\
 B &= -\alpha \Delta_t - \rho\sigma(y_{t+1} - \mu \Delta_t) \\
 C &= \alpha^2 \Delta_t^2 + 2\rho\sigma\alpha \Delta_t (\Delta y_{t+1} - \mu \Delta_t) + \sigma^2 (\Delta y_{t+1} - \mu \Delta_t)^2 - 2x_t^2 \alpha \sigma^2 (1 - \rho^2) \Delta_t.
 \end{aligned} \tag{2.20}$$

Les auteurs suggèrent comme valeur initiale de $\mathcal{L}_0(x_0) = e^{-x_0}$, où x_0 est la variance instantanée du modèle de Heston.

En procédant de manière récursive, il sera possible d'obtenir la valeur de x_1 pour ensuite déduire la valeur de $\mathcal{L}_1(x_1)$ sachant x_1 et $\mathcal{L}_0(x_0)$. Ce procédé récursif sera effectué jusqu'à ce que $t = T$ pour déduire la fonction de $p(x_T, y_{1:T})$ et donc, la fonction de vraisemblance. Par la suite, il suffit de résoudre le problème suivant :

$$\operatorname{argmax}_{\Omega \in \Omega} p(y_{1:T}) = \int p(x_T, y_{1:T}) dx_T \tag{2.21}$$

pour avoir les estimés de Ω .

Les conclusions de l'article indiquent que l'algorithme n'est pas efficient pour

estimer les paramètres, mais ce dernier filtre (prédit) la variance x_t précisément. Javeheri (2005) avait remarqué que la surface de la fonction de vraisemblance est aplatie et qu'il est difficile d'obtenir les paramètres du modèle par maximum de vraisemblance.

2.2.2.2 Aït-Sahalia & Kimmel (2007)

Les auteurs ont proposé une méthode d'estimation basée sur le processus joint bivarié du log-rendement et de la variance (y_t, x_t) . Dans le modèle de Heston, la variance est considérée latente alors, ils utilisent des pseudo-observations pour déterminer l'évolution du processus x_t . Un « *proxy* » analysé est la volatilité théorique du modèle de Black-Scholes $V_{\text{thec}}^{\text{BS}}$ extrait des prix observés sur les marchés financiers. Ils recommandent d'utiliser des produits dérivés « *at the money* » et de courte maturité pour mieux refléter la dynamique, car ce sont les options les plus liquides. Ainsi, ils vont utiliser une approximation décrite dans Aït-Sahalia (2001) pour déterminer la fonction de vraisemblance du processus joint et donc, être en mesure d'estimer les paramètres Ω du modèle de Heston en la maximisant. Cette méthode requiert une écriture du modèle sous la mesure de probabilité \mathbb{Q} , car ils utilisent les prix de produits dérivés dans leur méthode d'estimation et les paramètres résultant seront sous la mesure neutre au risque.

Également, ils mentionnent qu'utiliser la volatilité implicite de Black-Scholes peut produire un biais dans l'estimation des paramètres. Pour y remédier, ils utilisent un ajustement qui viendra éliminer ce biais et introduisent leur « *proxy* » (*the integrated volatility proxy*).

Supposons que x_t est la variance instantanée des log-rendements, il est possible

d'exprimer la moyenne de la variance intégrée sur la période $[t, T]$ comme

$$V(t, T) = \frac{1}{T-t} \int_t^T x_u du. \quad (2.22)$$

Cette équation est valide peu importe le modèle en question. Par la suite, leur « proxy » sera tout simplement de substituer la valeur de l'espérance de $V(t, T)$ par la volatilité implicite de Black-Scholes. En supposant une dérive d'un modèle quelconque de la forme $a + bx_t$, l'espérance de la valeur de $V(t, T)$ est

$$\mathbb{E}_t[V(t, T)] = \left(\frac{e^{b(T-t)} - 1}{b(T-t)} \right) \left(x_t + \frac{a}{b} \right) - \frac{a}{b}, \quad (2.23)$$

et leur « proxy » devient

$$x_t \approx \frac{bV_{\text{imp}}^{\text{BS}}(t, T) + a(T-t)}{e^{b(T-t)} - 1} - \frac{a}{b}. \quad (2.24)$$

Contrairement à l'équation (??), leur hypothèse concernant la prime de risque est

$$\lambda(S_t, x_t, t) = [\lambda_1^* \sqrt{(1-\rho^2)x_t}, \lambda_2^* \sqrt{x_t}]', \quad (2.25)$$

et le modèle de Heston résultant sous la mesure \mathbb{Q} devient

$$d \begin{bmatrix} \log S_t \\ x_t \end{bmatrix} = \begin{bmatrix} a + bx_t \\ \kappa^*(\theta^* - x_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)x_t} & \rho\sqrt{x_t} \\ 0 & \sigma\sqrt{x_t} \end{bmatrix} d \begin{bmatrix} \tilde{W}_t^{(1)} \\ \tilde{W}_t^{(2)} \end{bmatrix} \quad (2.26)$$

où

$$\kappa^* = \kappa - \lambda_2^* \sigma, \quad \theta^* = \left(\frac{\kappa + \lambda_2^* \sigma}{\kappa} \right) \theta, \quad a = r - q, \quad b = \lambda_1^*(1-\rho^2) + \lambda_2^* \rho - \frac{1}{2}. \quad (2.27)$$

Ainsi, ils obtiendront un estimateur de $\Omega = \{\kappa, \theta, \sigma, \rho, \lambda_1^*, \lambda_2^*\}$ dans le modèle de Heston sous la dynamique \mathbb{Q} .

2.2.2.3 Autres méthodes

Chernov & Ghysels (2000) ont appliqué une méthode des moments efficiente (EMM) basée sur les prix du sous-jacent en incluant des prix de produits dérivés pour déduire les paramètres Ω sous la mesure \mathbb{Q} . Également, des approches

bayésiennes ont été appliquées sur le modèle de Heston par Jones (2003) et Eraker (2004) en utilisant des prix du sous-jacent et d'options. Le modèle investigué par Eraker (2004) incorpore un fait stylisé supplémentaire qui est la présence de sauts. Christoffersen & al. (2010) ont utilisé des filtres à particules pour estimer divers modèles à volatilité stochastique dont celui de Heston et du SVOL. Ils utilisent une méthode par ML pour déterminer la dynamique sous \mathbb{P} mais également une autre méthode pour incorporer des prix d'options et déduire une dynamique neutre au risque.

2.3 Simulation du modèle de Heston

Dans le cas du modèle de Heston, il n'existe pas de solution explicite simple pour simuler de manière efficiente. Bref, une méthode de simulation approximative devra être envisagée. La première étape consistera à discrétiser le processus. En effet, la simulation doit être effectuée à différents temps discrets et ne peut être effectuée sur une base continue, car cela impliquerait une infinité d'observations. Une équation différentielle détermine le comportement local du processus sur un intervalle de longueur infinitésimale $[t, t + dt]$. Pour simuler le processus à chaque pas de temps Δ_t fini, il faut obtenir la solution de l'EDS, ce qui veut dire, l'intégrer sur un intervalle de temps $[t, t + \Delta_t]$. En supposant une EDS d'un processus H_t quelconque, la discrétisation de H_t revient à trouver la solution à

$$H_{t+\Delta_t} = H_t + \int_t^{t+\Delta_t} \mu(H_u, u) du + \int_t^{t+\Delta_t} \sigma(H_u, u) dW_u. \quad (2.28)$$

La simulation de H_t s'effectuera sur un intervalle de temps $[0, T]$ divisé en k points de sorte que $0 = t_0 < t_1 < \dots < t_k = T$ où chaque incrément de temps est uniformément dispersé avec une longueur de $\Delta_t = T/k$. Sachant la valeur de H_{t_0} et $W_{t_0} = 0$, il sera possible de déterminer la valeur de H_{t_1} selon la réalisation de W_{t_1} et procéder de manière itérative pour obtenir toutes les valeurs suivantes jusqu'à H_T .

Pour le processus de Heston, il faut appliquer la décomposition de Cholesky pour obtenir les deux réalisations de $W_{t+\Delta_t}^{(1)}$ et $W_{t+\Delta_t}^{(2)}$ de sorte que $\mathbb{E}[W_{t+\Delta_t}^{(1)} \cdot W_{t+\Delta_t}^{(2)}] = \rho$. Par la suite, le principe reste le même, il faut obtenir de manière itérative la valeur de la variable latente $x_{t+\Delta_t}$ à partir de $(x_t, W_{t+\Delta_t}^{(2)})$ et sachant la valeur de $x_{t+\Delta_t}$, obtenir la valeur de $y_{t+\Delta_t}$ selon la réalisation de $W_{t+\Delta_t}^{(1)}$.

Algorithme général

1. Initialiser le prix du sous-jacent S_0 (ou y_0 le log-rendement) ainsi que la variance instantanée x_0 .
2. Générer deux réalisations de normales indépendantes Z_1 et Z_2 puis appliquer la décomposition de Cholesky pour obtenir le couple

$$(Z_{t+\Delta_t}^{(1)} = Z_1, Z_{t+\Delta_t}^{(2)} = \rho Z_1 + \sqrt{1 - \rho^2} Z_2).$$

3. Calculer la réalisation de la variable latente $x_{t+\Delta_t}$
4. Sachant $x_{t+\Delta_t}$, calculer la réalisation de $S_{t+\Delta_t}$ (ou $y_{t+\Delta_t}$)
5. Aller à l'étape 2 jusqu'à ce que $t + \Delta_t = T$

Deux égalités sont nécessaires pour être en mesure d'appliquer cet algorithme qui sont :

$$\begin{aligned} W_{t+\Delta_t}^{(i)} - W_t^{(i)} &\stackrel{\mathcal{L}}{=} \sqrt{(t + \Delta_t - t)} Z_{t+\Delta_t}^{(i)} = \sqrt{\Delta_t} Z_{t+\Delta_t}^{(i)} \\ W_0^{(i)} &= 0 \end{aligned} \quad (2.29)$$

où Z_t sont des variables normales centrées et réduites et le symbole $\stackrel{\mathcal{L}}{=}$ signifie une égalité en loi.

Selon le schéma de discrétisation du modèle de Heston, il est possible de générer des valeur de x_t négative (variance négative) et ce même si la condition de Feller est respectée. En effet, cette condition est valide seulement pour les processus à

temps continu et la simulation est effectuée sur une base discrète.

Il y a deux solutions pour remédier à ce problème. La première est de tronquer la valeur de x_t (*full truncation scheme*) tandis que la deuxième est de prendre l'opposé de x_t (*reflection scheme*). Tronquer la valeur de x_t se résume par $x_t = \max(0, x_t)$ tandis que prendre l'opposé de x_t se résume par $x_t = |x_t|$ à chaque pas de temps.

Il n'y a pas de solution meilleure qu'une autre. Tronquer la valeur crée une variance de 0 ce qui n'est pas réaliste. D'un autre côté, prendre la valeur absolue peut changer le comportement du processus de variance. En effet, une grande valeur négative représente une période de faible volatilité et peut se transformer en une grande valeur positive qui représente une période de volatilité élevée.

2.3.1 Discrétisation selon les schémas d'Euler

Le schéma d'Euler est analogue à approximer une intégrale à partir d'un rectangle où la hauteur est associée au point gauche du processus. En reprenant l'équation (2.28), le terme en lien avec le coefficient de dérive est approximé par

$$\int_t^{t+\Delta_t} \mu(H_u, u) du \approx \int_t^{t+\Delta_t} \mu(H_t, t) du = \mu(H_t, t) \int_t^{t+\Delta_t} dt = \mu(H_t, t) \Delta_t. \quad (2.30)$$

Lors de la simulation, il est intuitif d'utiliser le point gauche, car au temps t la valeur de $\mu(H_t, t)$ est connue mais pas la valeur de $\mu(H_{t+\Delta_t}, t + \Delta_t)$. Le terme en lien avec le coefficient de diffusion utilise également ce principe et est approximé par

$$\int_t^{t+\Delta_t} \sigma(H_u, u) dW_u \approx \int_t^{t+\Delta_t} \sigma(H_t, t) dW_u \stackrel{\text{L}}{=} \sigma(H_t, t) \sqrt{\Delta_t} Z. \quad (2.31)$$

Alors, l'approximation avec les schémas d'Euler du processus H_t devient

$$\hat{H}_{t+\Delta_t} = \hat{H}_t + \mu(\hat{H}_t, t) \Delta_t + \sigma(\hat{H}_t, t) \sqrt{\Delta_t} Z \quad (2.32)$$

où Z est une loi normale de moyenne 0 et de variance unitaire. Il peut être démontré que la convergence sera assurée lorsque $\Delta_t \rightarrow 0$ et donc, $\hat{H}_{t+\Delta_t} \stackrel{c}{=} H_{t+\Delta_t}$.

En appliquant la discrétisation d'Euler aux EDS du modèle de Heston, le système devient

$$\begin{aligned} y_{t+\Delta_t} &= \log S_{t+\Delta_t} - \log S_t = \left(\mu - \frac{1}{2} x_t \right) \Delta_t + \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(1)} \\ x_{t+\Delta_t} - x_t &= \kappa(\theta - x_t) \Delta_t + \sigma \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(2)} \end{aligned} \quad (2.33)$$

et il est possible d'écrire les mêmes équations en remplaçant μ par r ainsi que $Z_{t+\Delta_t}^{(i)}$ par $\tilde{Z}_{t+\Delta_t}^{(i)}$ pour la version sous \mathbb{Q} .

2.3.2 Discrétisation selon les schémas de Milstein

Cette discrétisation sert à mieux calculer les coefficients de dérive ainsi que de diffusion d'un processus d'Itô. Pour assurer la convergence vers la solution exacte, la condition de Lipschitz doit s'appliquer et c'est le cas pour le modèle de Heston. Pour plus de détails concernant les dérivations, il est possible de se référer Glasserman (2003).

En appliquant la discrétisation de Milstein aux EDS du modèle de Heston, le système devient

$$\begin{aligned} y_{t+\Delta_t} &= \log S_{t+\Delta_t} - \log S_t = \left(\mu - \frac{1}{2} x_t \right) \Delta_t + \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(1)} \\ x_{t+\Delta_t} - x_t &= \kappa(\theta - x_t) \Delta_t + \sigma \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(2)} + \frac{1}{4} \sigma^2 \Delta_t \left((Z_{t+\Delta_t}^{(2)})^2 - 1 \right). \end{aligned} \quad (2.34)$$

La différence entre ce schéma et celui d'Euler est dans la dynamique de la variance où il y a un terme supplémentaire servant à réduire considérablement l'erreur de discrétisation. Par conséquent, la simulation de valeurs négatives sera moins fréquente qu'avec les schémas d'Euler. Pour obtenir le système sous \mathbb{Q} , il suffit de remplacer μ par r ainsi que $Z_{t+\Delta_t}^{(i)}$ par $\tilde{Z}_{t+\Delta_t}^{(i)}$ dans l'équation (2.34).

2.3.3 La méthode exacte

Broadie et Kaya (2006) démontrent la discrétisation exacte. Cependant, il est très difficile de simuler à partir de cette méthode, car elle requiert un haut niveau de calcul informatique et c'est pourquoi il existe d'autres schémas pour pallier à cette contrainte. Pour avoir plus de détails sur la méthode, il est également possible de se référer à Van Haastrecht et Pelsser (2010).

2.4 Méthode proposée pour l'estimation du modèle de Heston

Au meilleur de nos connaissances, il n'existe pas dans la littérature de méthodes d'estimation du modèle de Heston par maximum de vraisemblance n'utilisant que le prix d'action, i.e. une estimation similaire aux modèles SVOL ou GARCH. Les filtres à particules ont déjà été utilisés dans le contexte de ML, mais requiert un haut degré de calcul informatique. Les deux mêmes méthodes du chapitre 1 pourront être utilisées en effectuant une discrétisation du modèle de Heston.

Ces méthodes, le filtre à particule (PF) et le filtre non-linéaire discret (DNF), nécessitent des modèles à temps discret et ce n'est pas le cas du modèle de Heston. Cependant, ce modèle peut être discrétisé pour en avoir une écriture sous une version discrète.

En appliquant la discrétisation d'Euler au modèle, le système résultant est

$$\begin{aligned} \log S_{t+\Delta_t} - \log S_t &= \left(\mu - \frac{1}{2}x_t \right) \Delta_t + \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(1)} \\ x_{t+\Delta_t} - x_t &= \kappa(\theta - x_t) \Delta_t + \sigma \sqrt{x_t} \sqrt{\Delta_t} Z_{t+\Delta_t}^{(2)} \end{aligned} \quad (2.35)$$

et ce modèle converge vers le modèle de Heston (version continue) lorsque $\Delta_t \rightarrow 0$. Cette discrétisation permet d'avoir un modèle SVOL à temps discret écrit comme

étant :

$$\begin{aligned} y_{t+\Delta t} &= \left(\mu - \frac{1}{2}x_t \right) \Delta t + \sqrt{x_t} \sqrt{\Delta t} Z_{t+\Delta t}^S \\ x_{t+\Delta t} &= \kappa \theta \Delta t + (1 - \kappa \Delta t) x_t + \sigma \sqrt{x_t} \sqrt{\Delta t} Z_{t+\Delta t}^x. \end{aligned} \quad (2.36)$$

En posant $\kappa \theta \Delta t = \gamma$ et $1 - \kappa \Delta t = \phi$, l'écriture du modèle de Heston discrétisé sera similaire aux modèles SVOL du chapitre 1. Le modèle de Heston résultant est

$$\begin{aligned} y_t &= \lambda g(y_{t-1}, x_t, x_{t-1}) \times \Delta t + \sigma_{t-1} \sqrt{\Delta t} \epsilon_t \\ \sigma_t^2 &= x_t = \gamma + \phi x_{t-1} + \delta p(y_{t-1}, x_{t-1}) + \sigma_\nu \sqrt{x_{t-1} \Delta t} \nu_t \end{aligned} \quad (2.37)$$

où $\mathbb{E}[\epsilon_t \cdot \nu_t] = \rho$, $\lambda g(y_{t-1}, x_t, x_{t-1}) = \lambda - \frac{1}{2}x_{t-1}$, $\delta p(y_{t-1}, x_{t-1}) = 0$.

Pour appliquer les méthodes proposées (PF et DNF), les densités conditionnelles de y_t et x_t sont nécessaires pour approximer la vraisemblance. Ces densités sont déduites avec la décomposition de Cholesky se résumant par

$$\epsilon_t = \rho \nu_t + \sqrt{1 - \rho^2} \epsilon_t^*$$

où $\mathbb{E}[\epsilon_t^* \cdot \nu_t] = 0$ et l'équation de la variable latente s'écrivant

$$\nu_t = \frac{x_t - \gamma - \phi x_{t-1}}{\sigma_\nu \sqrt{x_{t-1} \Delta t}}.$$

Donc, le système se réécrit comme

$$\begin{aligned} y_t &= \lambda g(y_{t-1}, x_{t-1}) \times \Delta t + \sigma_{t-1} \sqrt{\Delta t} \rho \left(\frac{x_t - \gamma - \phi x_{t-1}}{\sigma_\nu \sqrt{x_{t-1} \Delta t}} \right) + \sigma_{t-1} \sqrt{\Delta t} \sqrt{1 - \rho^2} \epsilon_t^* \\ x_t &= \gamma + \phi x_{t-1} + \sigma_\nu \sqrt{x_{t-1} \Delta t} \nu_t \end{aligned} \quad (2.38)$$

et les densités conditionnelles (deux lois normales indépendantes) sont :

$$\begin{aligned} r(y_t | x_t, x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi x_{t-1} \Delta t (1 - \rho^2)}} \exp \left(-\frac{(y_t - \mu_t)^2}{2 x_{t-1} \Delta t (1 - \rho^2)} \right) \\ g(x_t | x_{t-1}, y_{t-1}) &= \frac{1}{\sqrt{2\pi \sigma_\nu^2 x_{t-1} \Delta t}} \exp \left(-\frac{(y_t - (\gamma + \phi x_{t-1}))^2}{2 \sigma_\nu^2 x_{t-1} \Delta t} \right) \end{aligned} \quad (2.39)$$

où

$$\mu_t = \lambda g(y_{t-1}, x_t, x_{t-1}) \times \Delta t + \frac{\rho}{\sigma_\nu} (x_t - \gamma - \phi x_{t-1}). \quad (2.40)$$

Ainsi, il sera possible d'estimer les coefficients du modèle de Heston en utilisant des bases de données à haute fréquence (journalière, intra-journalière, etc.), car l'objectif est que $\Delta_t \rightarrow 0$ pour assurer la convergence de la discrétisation.

2.4.1 Quelques ajustements aux méthodes proposées

Dans le modèle de Heston, la variable latente (i.e. : la variance) est définie sur les nombres réels positifs \mathbb{R}^+ alors que la variable latente du modèle SVOL est définie sur les nombres réels \mathbb{R} . Cela va nécessiter quelques ajustements aux méthodes proposées, car le domaine devient borné.

2.4.1.1 Filtre non-linéaire discret (DNF)

La borne inférieure sera strictement positive lors de la discrétisation de la variable latente. Donc, le vecteur \mathbf{X} représentant les N points uniformément dispersés doit être défini sur l'intervalle suivant :

$$\begin{aligned} \mathbf{X} &= [x^{(1)}, \dots, x^{(N)}] \\ x^{(1)} &= \max(0^+, E_0 - I_x), \quad x^{(N)} = E_0 + I_x \\ I_x &= \delta_x(N) \sqrt{V_0} \end{aligned} \quad (2.41)$$

engendrera une distance entre chaque point de

$$\Delta_x = \frac{E_0 + I_x - \max(0^+, E_0 - I_x)}{N - 1}.$$

Évidemment, il est supposé qu'il n'y a pas de masse sur les valeurs négatives de la variable latente. Alors, la construction des intervalles se fera dorénavant avec la borne inférieure suivante :

$$c(1) = 0. \quad (2.42)$$

Par conséquent, il faudra normaliser le vecteur U_t et la matrice q_t pour s'assurer que la somme des probabilités soit de 1 à chaque itération. En effet, la densité de la loi normale sera intégrée de 0 à ∞ en plusieurs parties (i.e. : $\int_0^\infty q(x^{(i)}|x^{(j)}, y_{t-1})$) et chaque partie devra être normalisée pour la masse omise entre $] -\infty, 0]$.

2.4.1.2 Filtre à particules

Lors de la simulation du modèle de Heston, le poids assigné à une particule ayant une réalisation négative (i.e. $x_t^{(i)} < 0$) sera nul (i.e. : $\omega_t^{(i)} = 0$). Le filtre à particules utilise un très grand nombre de simulations. Ce constat sera assurément observé et ce même avec des paramètres respectant la condition de Feller. Le filtre à particules proposé effectue un ré-échantillonnage et donc, la particule ayant une réalisation négative ne sera pas ré-échantillonnée. Elle sera remplacée par une particule ayant une variance positive avec un poids non nul.

2.5 Étude Monte Carlo

L'objectif de cette étude Monte Carlo sera de valider les méthodes et démarches employées pour approximer la vraisemblance du modèle de Heston un peu comme au Chapitre 1. Premièrement, la qualité de la fonction de vraisemblance approximée par le DNF sera vérifiée en plusieurs points aléatoires. La base de comparaison sera le filtre à particules avec un grand nombre de particules. Deuxièmement, les propriétés statistiques asymptotiques des estimateurs générés par le DNF seront analysées.

2.5.1 Étude de la fonction de vraisemblance

Encore une fois, l'approximation par filtre à particules sera considérée comme étant « exacte » lorsque N^* est grand, car plusieurs applications réalisées valident ce fait (i.e. : Del Moral (1996), Doucet & al. (2001), Doucet & Johansen (2008)

et Creal (2012)). Ainsi, l'approximation du DNF sera comparée par rapport à l'approximation du PF selon un jeu de paramètres et une série d'observations générée à partir de ce jeu. En effectuant plusieurs évaluations de la fonction de vraisemblance point par point, il sera possible de vérifier la présence d'un biais quelconque sur celle-ci.

L'étude se basera sur 50 000 jeux de paramètres tirés aléatoirement sur un domaine plausible de Ω . L'interprétation financière de Ω a déjà été abordée dans les sections plus haut. En ce sens, le tirage aléatoire des paramètres Ω s'effectuera sur un domaine précis dans le but de valider le DNF en économétrie financière. Évidemment, le jeu tiré aléatoirement doit satisfaire la condition de Feller. En considérant l'écriture du modèle de Heston dans sa version continue, les paramètres seront fonction du domaine suivant :

$$\begin{aligned}
 1\text{E-}6 &\leq \kappa \leq 10; \\
 0.01^2 &\leq \theta \leq 0.5^2; \\
 1\text{E-}10 &\leq \sigma_v \leq \sqrt{1.9999 \times \kappa \times \theta}; \\
 -0.99 &\leq \rho \leq 0.99; \\
 0.01 &\leq \mu \leq 0.1.
 \end{aligned} \tag{2.43}$$

Cela permet de représenter des processus volatiles (via un $\theta = 0.5^2$) tout en respectant la condition de Feller. En fixant un jeu Ω , il est possible de transformer les paramètres vers l'écriture du modèle de Heston discrétisée référencée par l'équation (2.38) pour appliquer les approximations du DNF et PF.

Dénotons par $\Omega^{(j)}$, $j = 1, \dots, 50000$, le jeu de paramètres tiré aléatoirement sur le domaine représenté par les équations (2.43). La trajectoire $y_{1:T}^{(j)}$ engendrée par ce jeu de paramètre est simulée selon les schémas de discrétisation de Milstein. Durant ce procédé, le problème de simulation de variance négative ($x_t < 0$) a

été résolu avec la solution de réflexion (i.e. : $x_t = |x_t|$) . Il est possible de générer 50 000 trajectoires du processus de Heston chacune ayant 500 observations journalières ($\Delta_t = 1/250$). Par la suite, il suffit d'évaluer chacune des approximations (DNF et PF) de la fonction de vraisemblance $\hat{f}(y_{1:500}, \Omega^{(j)})$ pour vérifier la présence de biais. Pour négliger les effets liés aux différentes approximations, le nombre de particules et la longueur du vecteur \mathbf{X} sont fixés à $N^* = 25000$ et $N = 250$. Néanmoins, il y aura toujours de légères différences, car ce sont des approximations numériques.

La figure 2.1 représente la correspondance entre la vraisemblance évaluée avec DNF (approximation) et avec PF (quasi-exacte). Ainsi, il y a une très forte correspondance sur l'ensemble des trajectoires générées et l'histogramme démontre une quasi-égalité numérique en distribution.

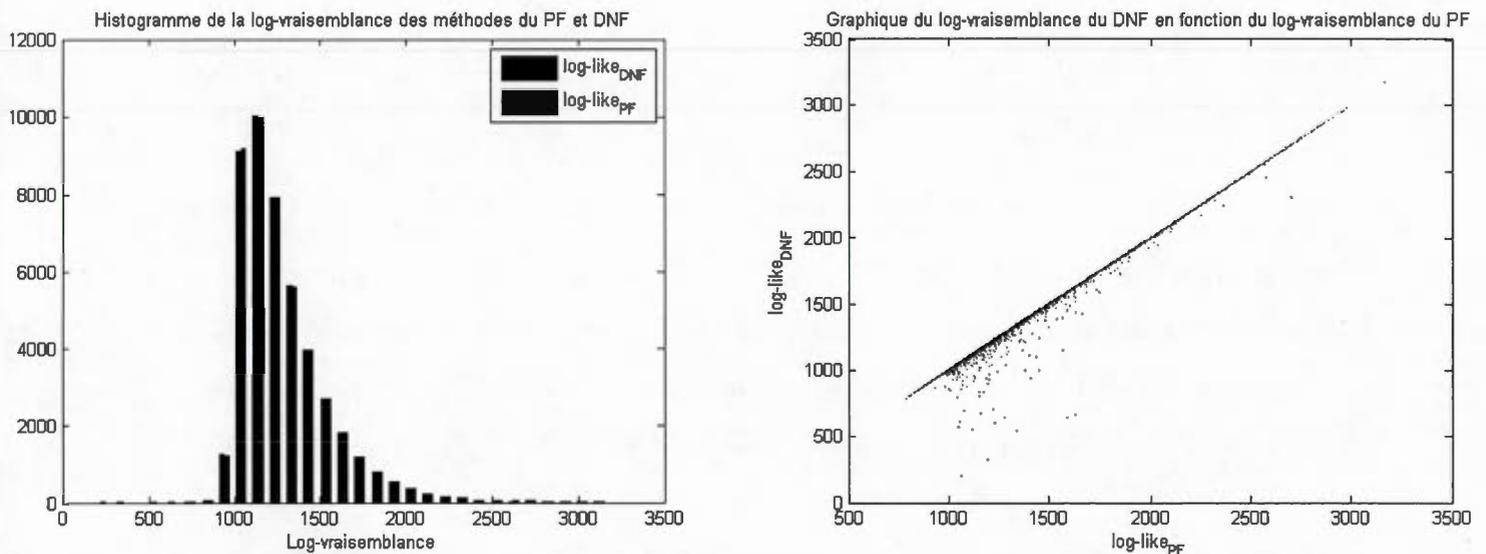


Figure 2.1: Histogramme et graphique des log-vraisemblances des deux méthodes DNF et PF selon 50 000 jeux de paramètres et trajectoires aléatoires.

Pour avoir une meilleure base de comparaison entre les différences des évaluations de $\hat{\ell} = \log\{f(\widehat{y}_{1:T}, \theta)\}$, il sera plus judicieux de travailler avec l'erreur relative de la méthode DNF, car elle permet une comparaison sur une base uni-dimensionnelle. L'erreur relative Δ_{rel} sera dénotée par

$$\Delta_{\text{rel}} = \frac{\hat{\ell}^{(\text{PF})} - \hat{\ell}^{(\text{DNF})}}{\hat{\ell}^{(\text{PF})}} * 100\%. \quad (2.44)$$

À partir de la figure 2.2, il est possible d'observer les mêmes constats élaborés avec le modèle SVOL. La grande majorité des erreurs se retrouve dans la zone $\pm 1\%$ et la dispersion des erreurs semblent être uniforme n'indiquant pas la présence de biais.

Il serait intéressant d'évaluer les quantiles empirique des erreurs. Pour ce faire,

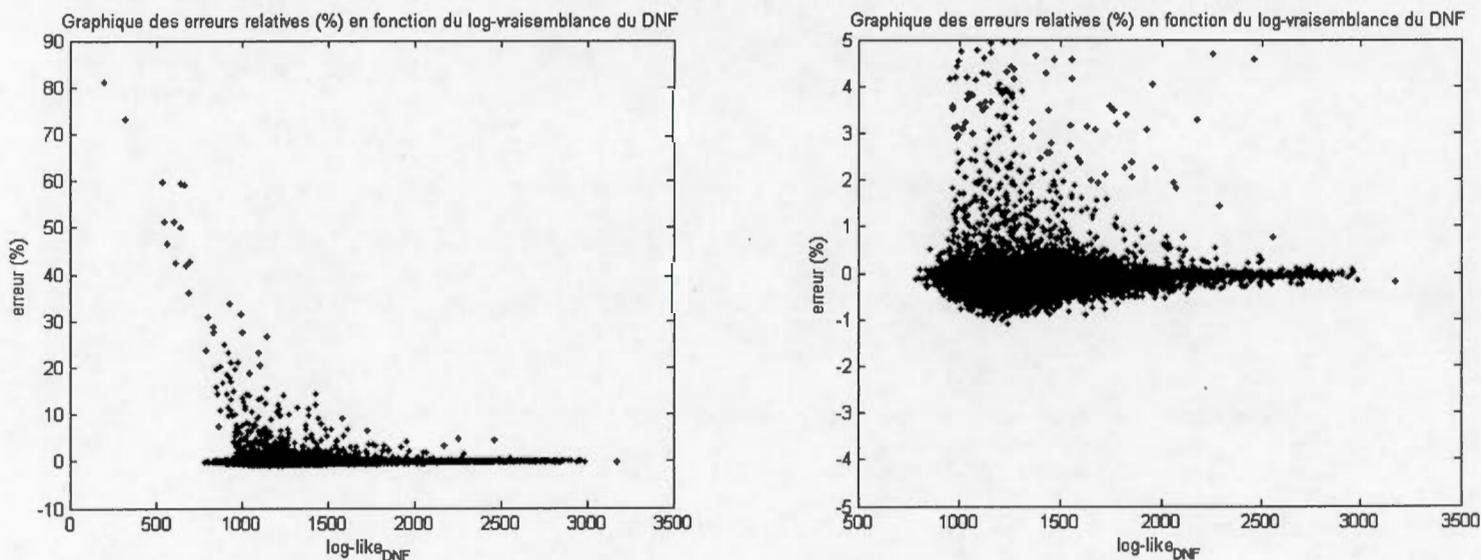


Figure 2.2: Graphique des erreurs relatives avec deux échelles d'axe en « y » différents selon les 50 000 jeux de paramètres et trajectoires aléatoires.

il vaut mieux ne pas tenir compte du signe, alors, dénotons les erreurs relatives

Niveau	Quantile
0.250	0.036347
0.500	0.080545
0.750	0.144425
0.900	0.243796
0.950	0.359963
0.990	0.815700
0.995	2.262125

Tableau 2.1: Quantiles empiriques des erreurs relatives absolues $\Delta_{\text{abs-rel}}$ (%) selon les 50 000 jeux de paramètres et trajectoires aléatoires.

absolues $\Delta_{\text{abs-rel}}$ comme

$$\Delta_{\text{abs-rel}} = |\Delta_{\text{rel}}|. \quad (2.45)$$

Les quantiles empiriques seront visualisés dans le Tableau 2.1. La méthode DNF commet une erreur de plus de 0.24% dans environ 10% des cas et une erreur de plus de 0.81% dans un peu plus de 1% des cas. En général, l'estimation avec DNF semble adéquate et l'approximation est sensiblement la même que celle du PF.

Les méthodes pour approximer la vraisemblance doivent être modifiées pour tenir compte du domaine de la variable latente qui est bornée positivement. Un peu comme dans le chapitre 1, il serait intéressant de cibler une certaine zone de paramètre où l'approximation du DNF est moins efficiente. En dénotant la condition de Feller comme étant :

$$\text{Feller} = 2\kappa\theta - \sigma_v^2 > 0, \quad (2.46)$$

il est possible d'analyser si la méthode DNF est moins efficiente dans la zone où la condition de Feller est respectée mais très proche de zéro.

Sur la figure 2.3, la méthode DNF évalue la vraisemblance moins précisément

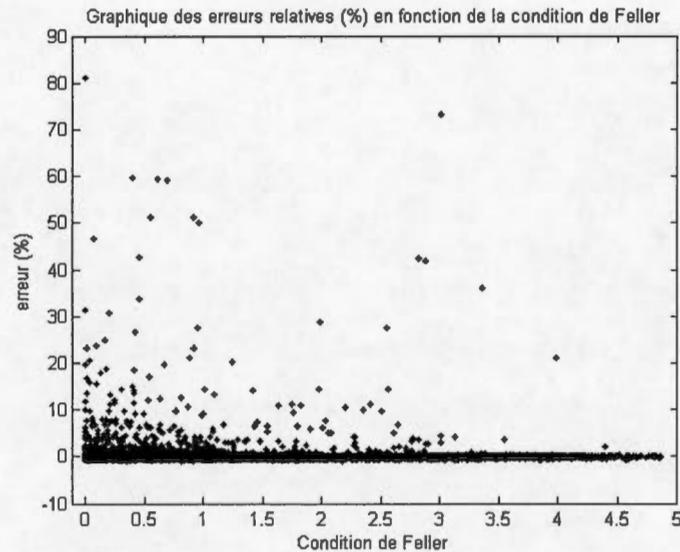


Figure 2.3: Graphique des erreurs relatives en fonction de la condition de Feller selon les 50 000 jeux de paramètres et trajectoires aléatoires.

lorsque la condition de Feller est proche de zéro. Néanmoins, les erreurs sont généralement près de zéro pour la grande majorité des cas ne justifiant pas le rejet de cette méthode dans cette zone de paramètre.

2.5.2 Étude Monte Carlo sur les propriétés statistiques du DNF

Dans la section 2.5.1, l'efficacité de la méthode DNF à approximer la vraisemblance pour une trajectoire donnée du modèle de Heston a été démontrée numériquement. Maintenant, le but sera de vérifier que le DNF estime bien les paramètres. En d'autres mots, déterminer si le maximum de vraisemblance a des propriétés statistiques désirables lorsqu'il est calculé par le DNF.

L'étude se basera sur le jeu de paramètres analysé dans l'article de Aït-Sahalia & Kimmel (2007) qui est :

$$\kappa = 3, \quad \theta = 0.1, \quad \sigma_\nu = 0.25, \quad \rho = -0.8.$$

Dans cet article, ils supposaient l'écriture du modèle sous une version dans la mesure de probabilités neutre au risque car ils utilisent les prix de produits dérivés dans leur méthode d'estimation. Ils supposaient une dérive en fonction d'un taux sans risque ($r = 0.04$) et d'un dividende ($q = 0.015$) fixe. Notre dérive sera composée d'un paramètre à estimer μ et il sera fixé lors des simulations à $\mu = 0.04$. À noter que cet article devrait estimer une prime de risque fixée lors des simulations à $\lambda_1^* = 4$ et considéraient $\lambda_2^* = 0$ (voir section 2.2.2.2).

Dans la même optique que l'article en question, il y aura 1000 trajectoires différentes de 500, 5000, 10000 observations quotidiennes ($\Delta_t = 1/252$) et 1000 trajectoires différentes de 500 observations hebdomadaires ($\Delta_t = 7/365$). Dans le tableau (2.2), le biais moyen et l'écart-type des estimations de chaque coefficient en fonction de la longueur de la trajectoire seront répertoriés. À noter que les simulations ont été effectuées avec des schémas de discrétisation de Milstein tout en utilisant la solution de réflexion lorsque $x_t < 0$.

Selon les résultats du Tableau 2.2, il est possible de conclure que la méthode DNF est sans biais et converge asymptotiquement vers Ω lorsque $T \rightarrow \infty$ pour le jeu de paramètres donné.

2.5.3 Conclusion

L'approximation de la fonction de vraisemblance avec la méthode DNF pour un point donné sur la courbe est correctement approximée et le maximum de cette courbe pour un jeu de donnée semble être bien centré. Cette méthode est recom-

	$\kappa = 3$	$\theta = 0.1$	$\sigma_\nu = 0.25$	$\rho = -0.8$	$\mu = 0.04$
$T = 500$	7.71418	-0.00130	0.06417	-0.05963	-0.00995
$(\Delta_t = 1/252)$	(25.3520)	(0.0170)	(0.2293)	(0.2474)	(0.2042)
$T = 5000$	0.12234	-2.0339E-4	1.3013E-4	-0.01162	-0.00358
$(\Delta_t = 1/252)$	(0.7113)	(0.0060)	(0.0326)	(0.0615)	(0.0673)
$T = 10000$	0.05535	-2.7394E-4	5.4550E-4	-0.00198	-0.00384
$(\Delta_t = 1/252)$	(0.4930)	(0.0043)	(0.0237)	(0.0440)	(0.0492)
$T = 500$	-1.6309	-9.6029E-4	-9.1952E-6	-0.0468	-0.0153
$(\Delta_t = 7/365)$	(0.8572)	(0.0245)	(0.0681)	(0.1488)	(0.0965)

Tableau 2.2: Résultats numériques de l'étude des propriétés asymptotiques du DNF dans le modèle de Heston. Le biais sur les paramètres et l'écart-type entre parenthèses y sont inscrit.

mandée pour estimer le modèle de Heston lorsque la fréquence d'observations du sous-jacent est quotidienne. Il serait intéressant de refaire l'exercice lorsque les données sont observées à des fréquences hebdomadaires ou mensuelles et analyser le comportement de la méthode. En d'autres mots, refaire l'expérience en simulation à très haute fréquence (i.e. : $\Delta_t = 1/1000$) puis redéfinir la base de données de sorte qu'elle représente des données, par exemple, mensuelles.

2.6 Exemple empirique

Cette section sert à comparer l'adéquation du modèle de Heston par rapport aux différentes versions du modèle à volatilité stochastique. Encore une fois, l'indice composé du S&P500 et l'indice composé du S&P/TSX¹ seront utilisées. Les deux

1. Les deux bases de données (S&P500 et S&P/TSX) sont obtenues à partir de Yahoo! Finance.

séries ont 1827 (respectivement 1826) observations hebdomadaires datant du 1er janvier 1980 au 1er janvier 2015.

Les modèles SVOL analysés seront ceux répertoriés dans l'analyse empirique du chapitre 1. Pour plus d'informations, se référer à la section 1.5.1.

Dans le Tableau 2.3, les critères d'information du modèle à volatilité stochastique de Heston sont plus faibles que ceux des autres versions. Ce qui veut dire que l'adéquation du modèle SVOL aux données des indices financiers est meilleure. En ce sens, le lien entre la variance et la variable latente semble mieux représenté dans le modèle SVOL (ou AAA) que celui de Heston. Dans le modèle SVOL, la variance en un temps donné est une log-normale tandis que dans le modèle de Heston, c'est une loi khi-deux non-centrée. À la lumière de ces résultats, la paramétrisation de la variance est mieux représentée par la distribution log-normale. Ceci est cohérent avec Christoffersen (2010) qui démontre que le modèle SVOL a une adéquation supérieure à celle de Heston dans le cadre de rendements d'indices boursiers.

À noter que les paramètres répertoriés dans le Tableau 2.3 sont sous une version transformée. Selon l'écriture usuelle du modèle de Heston, les paramètres (κ, θ) pour les indices respectivement du S&P/TSX et du S&P500 sont respectivement (3.04, 0.023) et (2.94, 0.024).

S&P/TSX	Heston	SVOL	SVOL (AAA)
γ / τ	1.3420E-3	-0.17786	0.13120
ϕ	0.94148	0.95612	0.04719
σ_ν	0.26678	0.24653	0.12559
ρ	-0.41662	-0.54364	-0.56977
λ/A	0.09605	0.12607	0.23561
B	x	x	-1.87103
C	x	x	6.53784
$\log f(y_{1:T}, \theta)$	4618.30	4649.04	4650.06
AIC	-9226.6	-9288.1	-9286.1
BIC	-9199.1	-9260.5	-9247.6
S&P500	Heston	SVOL	SVOL (AAA)
γ / τ	1.3691E-3	-0.15952	0.13915
ϕ	0.94348	0.95921	0.05560
σ_ν	0.25714	0.24200	0.12590
ρ	-0.54570	-0.63761	-0.71013
λ/A	0.10933	0.14529	-0.09517
B	x	x	2.24598
C	x	x	-2.21275
$\log f(y_{1:T}, \theta)$	4530.64	4558.66	4562.43
AIC	-9051.3	-9107.3	-9110.9
BIC	-9023.7	-9079.8	-9072.3

Tableau 2.3: Résultats du maximum de vraisemblance pour les trois modèles SVOL selon les indices composés du S&P/TSX et S&P500

CHAPITRE III

EFFICACITÉ DE STRATÉGIES DE COUVERTURE

3.1 Construction du marché dans un fonds distinct

Ces contrats offrent la flexibilité à un assuré d'investir dans les marchés financiers tout en ayant une protection contre les rendements défavorables de ceux-ci. La construction du portefeuille d'investissement de l'assuré dépend de son aversion au risque. Dans notre cas, ce portefeuille sera toujours composé à 100% en capital action. Ainsi, le générateur de scénarios économiques est composé d'un sous-jacent (actif) et d'un compte bancaire offrant un rendement fixe et sans risque de défaut.

3.1.1 Dynamique du sous-jacent

Le sous-jacent représente le prix d'un indice financier. Alors, il est une transformation de y_t qui est défini selon un modèle financier quelconque du chapitre 1 ou 2. Le prix au temps T est obtenu par

$$S_T = S_t \exp \left(\int_t^T y_u du \right) \quad (3.1)$$

où S_t est observé et il est supposé que $S_0 = 1$ pour des fins de simplification.

3.1.2 Dynamique de la valeur du fonds d'investissement

L'assuré paie une prime périodique notée α en échange de la garantie fournie par l'assureur. La prime vient diminuer la valeur du fonds donc, les contrats incorporent implicitement une certaine forme d'exonération des primes lorsque la garantie est « *in the money* ».

En pratique, la prime est payée à chaque intervalle de temps Δ_t , par exemple, à chaque mois ($\Delta_t = 1/12$). Suite au paiement de la prime, la valeur du fonds devient

$$A_t = A_{t-} e^{-\alpha \Delta_t} \quad (3.2)$$

où A_{t-} est la valeur du fonds tout juste avant le versement. Par la suite, le rendement du marché sera appliqué sur la valeur du fonds engendrant la relation suivante :

$$A_{t+\Delta t} = A_t \frac{S_{t+\Delta t}}{S_t}. \quad (3.3)$$

Dans la littérature, la prime est généralement payée sur une base continue pour réduire la complexité du problème lors de l'évaluation de la valeur de la garantie. La dynamique de la valeur du fonds est obtenue à l'aide du lemme d'Itô en posant $A_t = S_t e^{-\alpha t}$. L'EDS du fonds devient

$$dA_t = A_t \frac{dS_t}{S_t} - \alpha A_t dt \quad (3.4)$$

où dS_t réfère à l'EDS du sous-jacent. Par exemple, en supposant le sous-jacent obéissant à l'EDS de Black-Scholes, il en résulte la dynamique suivante :

$$\frac{dA_t}{A_t} = (\mu - \alpha)dt + \sigma dW_t. \quad (3.5)$$

Alors, l'EDS du fonds est en fait celle du sous-jacent payant un dividende continu de α .

3.1.3 Autres hypothèses de marchés

Dans la pratique, l'assureur prend différentes positions dans différents actifs du marché pour minimiser son exposition face aux risques générés par les fonds distincts. Pour simplifier le problème réel, le marché sera composé de seulement deux actifs, l'actif risqué (sous-jacent) et l'actif sans risque (compte bancaire).

Le compte bancaire offre un rendement sans risque de r sur une période de temps $[t, T]$. Il est défini par l'équation

$$B_T = B_t \exp(r(T - t)) \quad (3.6)$$

où B_t est l'investissement au temps t du compte bancaire.

De plus, le marché est supposé sans friction (aucun frais de transaction, aucun impôt, aucune taxe, etc), sans risque de liquidité et sans risque de défaut. Donc, les agents peuvent emprunter et investir au même taux sans risque de r . L'actif risqué sera transigé à un seul prix, il n'y a pas de différence entre le prix de vente et le prix d'achat de ce dernier (aucun « *bid-ask spread* »).

3.1.4 Le double risque du marché

Les primes perçues sont proportionnelles à la valeur du fonds d'investissement A_t . Lors de rendements défavorables du marché, la valeur du fonds d'investissement diminuera et engendre le double risque pour l'assureur. En effet, l'espérance du coût pour l'assureur augmentera et les primes perçues diminueront. L'effet global est doublement amplifié.

3.2 Les prestations garanties offertes par l'assureur

Il existe une multitude d'avenants permettant une flexibilité à l'assuré de définir son produit. Cette section sert d'en illustrer quelques-uns d'entre eux.

3.2.1 Prestation minimale garantie à l'échéance ou au décès

Deux cas très répandus sont les prestations minimales au décès et à maturité (GMDB et GMMB). Généralement, ces deux garanties viennent ensemble mais peuvent avoir des spécifications différentes. Ainsi, ces prestations s'appliquent sur la valeur du fonds au décès ou à la survie de l'assuré. Pour permettre l'évaluation de ces contrats, il faut introduire la notion d'échéance effective. C'est le premier événement entre l'échéance du contrat T et le temps du décès $T(x)$ d'un assuré

d'âge x à l'initialisation du contrat. L'échéance effective est dénotée par

$$T^* = \min(T(x), T). \quad (3.7)$$

Donc, le produit (incluant une GMDB et GMMB) offert par l'assureur est

$$\max(G_{T^*}, A_{T^*}) \quad (3.8)$$

où A_{T^*} est la valeur du fonds d'investissement à l'échéance effective et G_{T^*} est la valeur garantie du fonds. Ainsi, le produit assure une valeur minimale de G_{T^*} au client. Avec les propriétés du maximum, l'équation (3.8) se décompose comme

$$\max(G_{T^*}, A_{T^*}) = A_{T^*} + \max(0, G_{T^*} - A_{T^*}) \quad (3.9)$$

et le produit est tout simplement le retour du fonds d'investissement additionné d'une protection. La valeur du fonds à l'échéance effective n'est pas un risque pour l'assureur, donc, le coût de la garantie à l'échéance effective C_{T^*} est

$$C_{T^*} = \begin{cases} G_{T^*} - A_{T^*} & A_{T^*} < G_{T^*} \\ 0 & A_{T^*} \geq G_{T^*} \end{cases} \quad (3.10)$$

et ressemble au paiement d'une option de vente.

3.2.2 Dynamique de la valeur garantie G_t

La plupart des contrats incorporent plusieurs avenants pour définir la dynamique de la valeur garantie. Par exemple, un contrat GMMB pourrait offrir une garantie fixée à 75% de l'investissement initial avec un rendement minimal de 2% par an et un ajustement automatique aux rendement des marchés à tous les 6 mois. La dynamique de la valeur garantie varie d'un assureur à l'autre dans le but d'offrir un produit compétitif et attrayant.

3.2.2.1 Valeur de la garantie fixe

Le cas le plus simple est une garantie constante proportionnelle à l'investissement initial et se note

$$G_{T^*} = cA_0 \quad (3.11)$$

où A_0 est la valeur initiale du fonds et c est une proportion (i.e. : 75% ou 100%).

3.2.2.2 Rendement minimal sur la valeur de la garantie

Il est possible d'offrir un rendement minimal continument composé de $100b\%$ sur la valeur garantie. Supposons que la prime est prélevée à chaque intervalle de temps Δ_t , la relation entre la valeur garantie au temps $t - \Delta_t$ et t est

$$G_t = G_{t-\Delta_t} e^{b\Delta_t}. \quad (3.12)$$

3.2.2.3 Ajustement de la valeur de la garantie en fonction des rendements du marché

Une autre alternative est de fixer la valeur de la garantie en fonction de la dynamique du fonds d'investissement. En supposant une partition $M = 1, \dots, m$ où $t^{(1)} < \dots < t^{(m)}$ sont des temps déterminés d'avance à intervalle fixe (par exemple : à chaque anniversaire de police), le calcul effectué pour déterminer la valeur de la garantie au temps $t^{(k)}$ est

$$G_{t^{(k)}} = \max(G_{t^{(k)}-\Delta_t} e^{b\Delta_t}, cA_{t^{(k)}}). \quad (3.13)$$

Cette garantie est une sorte d'option rétroviseur.

Une autre manière d'ajuster la valeur de la garantie aux rendements des marchés est de laisser la liberté à l'assuré de spécifier les moments où la garantie sera réinitialisée et donc, fixée. En supposant une partition de moments $M = 1, \dots, m$

où $\tau^{(1)} < \dots < \tau^{(m)}$ sont des temps aléatoires déterminés par l'assuré, la relation de la valeur de garantie au temps $\tau^{(m)}$ est

$$G_{\tau^{(m)}} = cA_{\tau^{(m)}} \quad (3.14)$$

où $cA_{\tau^{(m)}} > G_{\tau^{(m-1)}}$. Évidemment, il est supposé que l'assuré fera des ré-initialisations avantageant son produit alors, la valeur de la garantie correspond au dernier temps aléatoire. En analogie aux produits dérivés, cette garantie ressemble aux options cliquets.

3.3 Efficacité des stratégies de couverture

L'objectif est d'illustrer les effets du risque de marché sur la couverture de produits simples. Ces produits sont choisis pour éviter de calculer les positions à l'aide de simulations Monte Carlo. En effet, les contrats ayant plusieurs avenants spécifiques n'ont pas de formules fermées pour évaluer les coûts des protections. Dans ces cas, la stratégie de couverture doit être évaluée par simulation Monte Carlo à chaque rebalancement du portefeuille de réplication et après chaque simulation du sous-jacent dans la mesure de probabilité réelle. Bref, cela engendre un problème calculatoire très laborieux. Donc, l'idée est de simplifier quelques hypothèses pour arriver à un produit s'évaluant facilement.

Lorsqu'un assureur a un portefeuille avec suffisamment d'assurés et diversifié, les hypothèses concernant les mathématiques actuarielles peuvent être assumées comme étant déterministe. Ce constat s'applique généralement pour la plupart des assureurs. En considérant la mortalité comme étant diversifiable et indépendante du coût des produits, la valeur présente d'une garantie $V(\text{GM} \cdot \text{B}, t)$ (i.e. : GMMB ou GMDB) au temps t s'évalue en conditionnant sur le décrement de mortalité. Alors, cela se résume par une double espérance notée

$$V(\text{GM} \cdot \text{B}, t) = \mathbb{E}_t^{\mathbf{P}} \left[\mathbb{E}_t^{\mathbf{Q}} \left[\frac{B_t}{B_{T^*}} C_{T^*} | T(x) \right] \right] \quad (3.15)$$

où $\mathbb{E}_t[\cdot]$ signifie l'espérance d'une variable aléatoire conditionnelle à toute l'information disponible jusqu'au temps t inclusivement. La première espérance $\mathbb{E}_t^{\mathbb{P}}[\cdot]$ au temps t s'évalue en prenant la mortalité, sous la mesure de probabilité physique, sur une base déterministe. Cela veut dire pondérer les toutes les possibilités que l'assuré survive jusqu'à $T(x)$ (respectivement $T(x) - 1$ et décède dans la période suivante) pour la garantie GMMB (respectivement GMDB). La deuxième espérance évalue le coût avec les notions d'ingénierie financière sachant le temps de décès de l'individu donc, l'échéance effective.

Par conséquent, un produit s'évaluant facilement est la garantie GMMB où la valeur de la garantie est fixée à l'initialisation et s'écrit

$$V(\text{GMMB}, t) = {}_{T-t}p_x \times \mathbb{E}_t^{\mathbb{Q}} \left[\frac{B_t}{B_T} C_T | T(x) \right]. \quad (3.16)$$

Il est possible d'omettre le décrétement de mortalité ${}_{T-t}p_x$ car ce terme est généralement près de 100% pour les assurés d'âge inférieur à 65 ans et pour une échéance de moins de 15 ans. Avec ces hypothèses, la valeur présente de ce produit est

$$V^*(\text{GMMB}, t) \approx \mathbb{E}_t^{\mathbb{Q}} \left[\frac{B_t}{B_T} \max(0, cA_0 - A_T) \right] \quad (3.17)$$

où V^* signifie la valeur présente en ne considérant pas la mortalité. Alors, ce produit est similaire à une option de vente sur la valeur du fonds d'investissement avec un prix d'exercice égale à une fraction de l'investissement initial et ayant une échéance de T ans.

3.3.1 Modèle sous la mesure de probabilité neutre au risque \mathbb{Q}

Il est nécessaire de spécifier un modèle sous la mesure de probabilité \mathbb{Q} à partir duquel l'évaluation de produits dérivés s'effectue facilement (i.e : les garanties offertes par l'assureur). En effet, l'idée est d'analyser l'efficacité d'une stratégie de couverture delta et donc, calculer le delta du modèle sous \mathbb{Q} à chaque pas de

temps pour chaque simulation de S_t selon chacun des modèles de marché choisis. Il y a deux choix possibles : le modèle de Black-Scholes et le modèle de Heston.

Afin d'évaluer l'impact du risque de marché sur l'efficacité de stratégies de couverture, celle-ci sera effectuée avec un modèle plus simple que le supposé modèle de marché. Un modèle couramment utilisé pour la couverture de fonds distincts est le modèle de Black-Scholes et son EDS sous la mesure de probabilité \mathbb{Q} est dénotée par

$$dS_t = S_t r dt + S_t \sigma^* d\tilde{W}_t \quad (3.18)$$

où r est le taux sans risque et σ^* est la volatilité implicite attendu à long terme du modèle sous la mesure \mathbb{P} (voir équation (3.25)). À noter que dans la pratique, il n'est pas rare de retrouver ce choix de modèle dans les hypothèses des assureurs.

3.3.2 Couverture delta d'une garantie minimale à l'échéance (GMMB) avec une prime unique

Pour se protéger du double risque du marché, l'assureur peut charger une prime unique à l'initialisation du contrat. La prime unique π diminue la valeur du fonds à l'initialisation tel que

$$A_0^+ = A_0 - \pi \quad (3.19)$$

et l'évolution de l'investissement devient

$$A_t = A_0^+ \times S_t. \quad (3.20)$$

Ainsi, la juste prime de ce produit est noté par

$$\pi = V^*(GMMB, t = 0) = P(A_0^+, K = cA_0, T) \quad (3.21)$$

où le produit offert par l'assureur est une option de vente $P(\cdot)$ d'échéance T avec un prix d'exercice de $G_{T^*} = cA_0$ et une valeur initiale du fonds de A_0^+ . La juste

prime sera trouvée en résolvant l'égalité suivante

$$\pi - P(A_0 - \pi, K = G_{T^*}, T) = 0 \quad (3.22)$$

avec une méthode numérique de recherche du zéro d'une fonction.

3.3.2.1 Hypothèses et construction du portefeuille de réplication

Le risque de modèle de marché sera analysé avec les modèles BS, GARCH, RSLN, MJD, SVOL, AAA et Heston où les paramètres sont répertoriés dans les tableaux 1.7 et 2.3.

Une couverture delta implique avoir une position dans l'actif risqué S_t égale au delta Δ_t^{BS} calculé avec le modèle sous la mesure \mathbb{Q} (i.e. : Black-Scholes). Cette position est

$$\frac{\partial P(A_t, G_{T^*}, T - t)}{\partial A_t} = \Delta_t^{\text{BS}} = -N(-d_1) = - \int_{-\infty}^{-d_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (3.23)$$

où

$$d_1 = \frac{\log\left(\frac{A_t}{G_{T^*}}\right) + \left(r + \frac{1}{2}\sigma^{*2}\right)(T - t)}{\sigma^* \sqrt{T - t}}. \quad (3.24)$$

Le paramètre σ^* sera fixé à la volatilité implicite à long terme attendu du modèle utilisé sous la mesure de probabilité \mathbb{P} comme étant :

$$\text{BS : } \sigma^* = \sigma_1$$

$$\text{GARCH : } \sigma^* = \sqrt{\frac{\omega}{1 - \alpha - \beta}}$$

$$\text{RSLN : } \sigma^* = \sqrt{\pi_1 \sigma_1^2 + \pi_2 \sigma_2^2}$$

$$\text{MJD : } \sigma^* = \sqrt{\alpha (\log(1 + \beta) - 0.5\sigma_2^2)^2 + \sigma_1^2} \quad (3.25)$$

$$\text{SVOL : } \sigma^* = \sqrt{\exp\left\{\frac{\gamma}{(1 - \phi)} + \frac{\sigma^2}{2(1 - \phi^2)}\right\}}$$

$$\text{Heston : } \sigma^* = \sqrt{\frac{\gamma}{1 - \phi}}$$

$$\text{AAA : } \sigma^* \text{ est approximé par simulation Monte Carlo.}$$

En dénotant la valeur du portefeuille de réplication Π_t (respectivement Π_{t-}) suite au re-balancement (respectivement avant le re-balancement), la stratégie de couverture delta autofinancée se résume par

$$\begin{aligned}\Pi_{t-} &= \Delta_{t-\Delta t}^{\text{BS}} S_t + \Psi_{t-\Delta t} B_t \\ \Pi_t &= \Delta_t^{\text{BS}} S_t + \Psi_t B_t\end{aligned}\tag{3.26}$$

où Ψ_t est l'investissement dans le compte bancaire. Cette position est déduite de sorte que

$$\Pi_t = \Pi_{t-}\tag{3.27}$$

où $\Pi_0 = \pi$. Alors, l'erreur de réplication à l'échéance se calcule comme

$$\text{Err}_{(T)} = C_T - \Pi_T.\tag{3.28}$$

3.3.2.2 Stratégie de couverture nulle (perte non-couverte)

La perte non-couverte de l'assureur suppose une position nulle dans le sous-jacent

$$\Delta_t^{\text{BS}} = 0\tag{3.29}$$

et l'erreur de réplication à l'échéance de cette stratégie sera calculée comme étant :

$$\text{Err}_{(T)} = C_T - \frac{B_T}{B_t} \pi.\tag{3.30}$$

Donc, cette stratégie implique aucun acte de mitigation concernant le modèle de marché. L'assureur est à l'aise avec la tarification et la distribution de la perte engendrée par son produit. Cette méthode est l'approche actuarielle et consiste à mettre un montant d'argent de côté de sorte que la compagnie d'assurance remplit ses obligations à un certain seuil. Ce montant est généralement calculé avec une mesure de risque comme la valeur au risque (VaR) ou l'espérance conditionnelle de queue (ECQ ou CTE).

3.3.2.3 Algorithme

La première étape consiste à choisir la stratégie de couverture (nulle ou delta), le modèle de marché (BS, GARCH, RSLN, MJD, SVOL, Heston, AAA), les paramètres de l'indice du S&P500 ou du S&P/TSX et la fréquence de rebalancement du portefeuille de réplcation Δ_t . Par la suite, l'algorithme ci-dessous est appliqué.

Initialisation ($t = 0$)

- Calcul de la volatilité σ^* utilisé sous la mesure de probabilité \mathbb{Q} selon le modèle de marché sélectionné.
- Calcul de la prime unique π et de la position dans le sous-jacent Δ_0^{BS} selon la stratégie utilisée.
- Poser $\Pi_0 = \pi$ et déduire l'investissement dans la compte bancaire Ψ_0 .

Boucle ($t = \Delta_t, \dots, T - \Delta_t$)

- Simulation du sous-jacent S_t sous la mesure de probabilité réelle selon le modèle de marché sélectionné et déduire la valeur du fonds A_t .
- Calcul de la valeur du portefeuille avant le rebalancement

$$\Pi_{t-} = \Delta_{t-\Delta_t}^{\text{BS}} S_t + \Psi_{t-\Delta_t} B_t.$$

- Calcul du Δ_t^{BS} selon la stratégie utilisée.
- Poser $\Pi_{t-} = \Pi_t$ et déduire l'investissement dans la compte bancaire Ψ_t .

Étape finale ($t = T$)

- Simulation du sous-jacent S_T selon le modèle de marché sélectionné et déduire la valeur du fonds A_T .
- Calcul de la valeur final du portefeuille

$$\Pi_T = \Delta_{T-\Delta_t}^{\text{BS}} S_T + \Psi_{T-\Delta_t} B_T.$$

- Calcul de l'erreur de réplcation

$$\text{Err}_{(T)} = C_T - \Pi_T.$$

	Moyenne	Écart-type	Corr($C_T; \Pi_T$)	CTE(95%)	VaR(99%)
BS	-1.73 / -4.69	11.68 / 7.77	0.00 / 0.00	39.27 / 25.84	46.90 / 35.50
GARCH	-7.10 / -7.83	5.54 / 3.61	0.00 / 0.00	11.55 / 0.23	24.63 / 9.45
RSLN	11.68 / 4.31	23.21 / 18.90	0.00 / 0.00	67.07 / 58.26	72.43 / 64.92
MJD	-3.32 / -5.43	8.78 / 5.59	0.00 / 0.00	30.34 / 15.21	39.02 / 26.35
SVOL	21.94 / 28.75	25.32 / 27.44	0.00 / 0.00	71.57 / 76.51	75.47 / 79.65
Heston	-4.38 / -5.11	4.60 / 3.76	0.00 / 0.00	10.63 / 4.78	22.15 / 15.24
AAA	26.00 / 22.50	25.18 / 24.70	0.00 / 0.00	72.45 / 70.28	76.13 / 74.27

Tableau 3.1: Résultats de la perte non-couverte de la garantie GMMB. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.

3.3.3 Résultats numériques

L'expérience est réalisée avec les spécifications suivantes sur le GMMB :

$$A_0 = 100, \quad G_{T^*} = 100\%A_0, \quad T = 10, \quad r = 0.05,$$

et en générant 100 000 scénarios sous la mesure \mathbb{P} de différents modèles de marché. La fréquence des re-balancements des positions se fera sur une base mensuelle ($\Delta_t = 1/12$), hebdomadaire ($\Delta_t = 1/52$) et journalière ($\Delta_t = 1/252$).

À partir des tableaux 3.1, 3.2, 3.3 et 3.4, le risque de modèle de marché est bel et bien présent. En effet, la distribution de la perte lorsqu'il n'y a aucune stratégie de couverture varie grandement en fonction du modèle de marché. Les modèles générant les queues de distribution les plus lourdes sont les modèles SVOL, AAA et RSLN.

	Moyenne	Écart-type	Corr($C_T; \Pi_T$)	CTE(95%)	VaR(99%)
BS	0.00 / 0.02	1.34 / 1.20	99.34 / 98.81	3.24 / 2.92	3.87 / 3.52
GARCH	0.07 / 0.11	2.89 / 2.44	87.03 / 79.63	7.90 / 6.83	9.60 / 8.32
RSLN	0.47 / 0.55	3.24 / 2.67	99.10 / 99.06	8.29 / 7.40	9.56 / 8.59
MJD	-0.01 / 0.02	1.45 / 1.25	98.67 / 97.58	3.48 / 3.06	4.17 / 3.71
SVOL	0.76 / 0.93	2.64 / 2.89	99.53 / 99.53	7.13 / 7.76	8.21 / 8.90
Heston	0.12 / 0.16	0.98 / 0.94	97.73 / 96.84	2.80 / 2.80	3.42 / 3.42
AAA	0.38 / 0.25	2.34 / 2.54	99.63 / 99.56	5.72 / 6.14	6.52 / 7.03

Tableau 3.2: Résultats des erreurs de réplifications $\text{Err}(T)$ de la garantie GMMB où la fréquence de re-balancement est mensuelle. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.

	Moyenne	Écart-type	Corr($C_T; \Pi_T$)	CTE(95%)	VaR(99%)
BS	0.00 / 0.00	0.65 / 0.58	99.84 / 99.72	1.57 / 1.40	1.88 / 1.67
GARCH	0.01 / 0.03	2.14 / 1.86	92.56 / 87.24	6.28 / 5.41	7.77 / 6.76
RSLN	0.15 / 0.23	2.58 / 2.00	99.43 / 99.47	6.18 / 5.22	7.15 / 6.11
MJD	-0.02 / -0.01	0.99 / 0.82	99.38 / 98.94	2.34 / 1.98	2.81 / 2.39
SVOL	0.27 / 0.24	1.65 / 1.80	99.81 / 99.82	4.00 / 4.12	4.60 / 4.72
Heston	0.09 / 0.12	0.54 / 0.52	99.32 / 99.05	1.59 / 1.60	1.93 / 1.96
AAA	0.08 / 0.23	1.54 / 1.52	99.84 / 99.84	3.41 / 3.55	3.92 / 4.07

Tableau 3.3: Résultats des erreurs de réplifications $\text{Err}(T)$ de la garantie GMMB où la fréquence de re-balancement est hebdomadaire. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.

	Moyenne	Écart-type	Corr($C_T; \Pi_T$)	CTE(95%)	VaR(99%)
BS	0.00 / 0.00	0.30 / 0.27	99.97 / 99.94	0.71 / 0.64	0.86 / 0.77
GARCH	0.01 / 0.01	2.03 / 1.74	93.33 / 88.67	5.89 / 5.03	7.18 / 6.24
RSLN	0.06 / 0.15	2.40 / 1.82	99.50 / 99.56	5.63 / 4.61	6.56 / 5.37
MJD	-0.02 / -0.01	0.84 / 0.68	99.55 / 99.27	1.95 / 1.63	2.37 / 1.98
SVOL	0.11 / 0.03	1.31 / 1.42	99.88 / 99.89	2.92 / 2.86	3.33 / 3.27
Heston	0.09 / 0.11	0.34 / 0.33	99.72 / 99.62	1.07 / 1.09	1.31 / 1.32
AAA	0.00 / 0.23	1.24 / 1.11	99.89 / 99.91	2.58 / 2.53	2.96 / 2.87

Tableau 3.4: Résultats des erreurs de réplifications $Err_{(T)}$ de la garantie GMMB où la fréquence de re-balancement est journalière. La colonne de gauche représente les résultats selon les paramètres du S&P/TSX tandis que celle de droite est selon l'indice du S&P500.

S&P/TSX	BS	GARCH	RSLN	MJD	SVOL	Heston	AAA
π	3.84	4.90	3.85	3.64	3.52	3.13	2.60
σ^*	16.81%	16.19%	16.26%	15.90%	15.73 %	15.14%	14.29%
S&P500	BS	GARCH	RSLN	MJD	SVOL	Heston	AAA
π	4.10	5.01	3.98	3.97	4.44	3.41	3.02
σ^*	16.54%	17.75%	16.37%	16.36%	17.00 %	15.56%	14.97%

Tableau 3.5: La prime unique π et la volatilité implicite utilisée dans le modèle de Black-Scholes sous la mesure \mathbb{Q} lors des résultats des erreurs de couverture du GMMB.

Selon les résultats, en utilisant une stratégie de couverture delta tout simplement, il est possible de pratiquement éliminer le risque de modèle de marché. En effet, l'efficacité de la stratégie de couverture est vérifiée avec la diminution des trois mesures de risque (l'écart-type, la CTE(95%) et la VaR(99%)) sur la distribution de la perte. Ceci est encore plus marqué lorsque la fréquence de re-balancement augmente. Cela est bien résumé par la corrélation entre le portefeuille de réplique Π_T et le coût C_T qui augmente signifiant que la stratégie de couverture réplique dans la très grande majorité du temps le produit.

À notre grande surprise, les modèles GARCH et RSLN sont ceux où l'amélioration de l'efficacité de la couverture est la moins marquée parmi les autres modèles. Une hypothèse expliquant ce constat peut être la paramétrisation de ces modèles générant les plus grandes volatilités à long terme (voir Tableau 3.5). En ce sens, il est possible de conclure que le choix du modèle de marché va influencer l'hypothèse de volatilité à long terme attendue sous la mesure de probabilité réelle et donc, la tarification et les positions dans le sous-jacent à travers le temps. Il est bien de rappeler que tous les modèles ont été estimés sur les mêmes séries de rendements illustrant le risque de modèle de marché dans les fonds distincts.

Évidemment, il y a des limites à nos conclusions. L'expérience a été réalisée avec des hypothèses idéales (aucun frais de transactions, aucune différence entre le prix d'achat et de vente du sous-jacent et des obligations zéro-coupon, etc). De plus, les produits sont plus complexes dans leur évaluations (avenant sur la valeur garantie, décaissement du fonds, déchéance dynamique, taux d'intérêt stochastique, mortalité, allocation du fonds dans plusieurs actifs, prime périodique, etc) et une couverture delta n'aura pas autant de bénéfice. Cependant, il devrait en résulter que le risque lié à la perte sera diminué.

CONCLUSION

Dans ce mémoire, nous avons analysé l'estimation par maximum de vraisemblance de divers modèles à volatilité stochastique (temps discret et continu). Pour y parvenir, nous avons comparé les méthodes de MacDonald & Zucchini (1997) et des filtre à particules. Nous avons constaté que la méthode du filtre non-linéaire discret est rapide et sans biais. De plus, nous avons étendu l'approche par chaînes de Markov cachées pour l'estimation du modèle de Heston. Nous avons également observé que le modèle à volatilité stochastique à temps discret a une meilleure adéquation que le modèle de Heston. Finalement, nous avons appliqué des stratégies de couverture delta neutre dans un contexte de fonds distincts où le modèle de marché est inconnu.

Dans un avenir proche, nous aimerions valider si l'estimation avec notre adaptation pour le modèle de Heston est toujours valide lorsque la fréquence d'observation des rendements est discrète (hebdomadaire, mensuelle ou annuel). De plus, nous souhaitons évaluer l'efficacité d'une stratégie de couverture delta neutre mais basée sur le modèle de Heston. L'objectif serait de déterminer si nous améliorons beaucoup l'efficacité de la couverture en utilisant une hypothèse plus réaliste par rapport au modèle de Black-Scholes.

APPENDICE A

RÉSULTATS DU MAXIMUM DE VRAISEMBLANCE

Les paramètres du maximum de vraisemblance de différents modèles (Black-Scholes, GARCH (1,1), RSLN (2 états) et Merton-Jump Diffusion) sur les indices financiers du S&P/TSX et S&P500 apparaissent dans le tableau A.1.

Cependant, les paramètres du maximum de vraisemblance pour le Markov-switching GARCH(1,1) n'est pas dans le même format. En effet, ce modèle est plus difficile à estimer et à converger alors, il vaut mieux travailler avec des rendements en pourcentage (i.e. : $y'_t = y_t * 100$). De plus, les paramètres ne sont pas sur une base annuelle. Ainsi, les paramètres pour ce modèle sont dans le tableau A.2. Toutefois, la valeur de la vraisemblance a été ajustée pour être comparée aux autres modèles.

S&P/TSX	BS	GARCH(1,1)	RSLN(2 états)	MJD
γ_1	x	1.1362E-3	x	x
α	x	0.16113	x	12.7341
β	x	0.80222	x	6.8961E-3
σ_1	0.16177	x	0.11127	0.10589
σ_2	x	x	0.28551	0.0326
p_{11}	x	x	0.41924	x
p_{22}	x	x	0.03042	x
μ_1	0.06007	0.11256	0.12086	0.07251
μ_2	x	x	-0.12437	x
S&P500	BS	GARCH(1,1)	RSLN(2 états)	MJD
γ_1	x	1.1287E-3	x	x
α	x	0.15376	x	9.66914
β	x	0.81040	x	-5.4678E-3
σ_1	0.16537	x	0.11463	0.12211
σ_2	x	x	0.26289	0.03466
p_{11}	x	x	0.26849	x
p_{22}	x	x	0.016713	x
μ_1	0.08465	0.13388	0.16082	0.09779
μ_2	x	x	-0.09206	x

Tableau A.1: Résultats du maximum de vraisemblance pour les quatre modèles stochastiques selon les indices composés du S&P/TSX et S&P500.

Indice	γ_1	γ_2	α	β	p_{12}	p_{21}	μ_1	μ_2
S&P/TSX	0.07750	1.00344	0.05101	0.90345	0.04939	0.46043	0.27837	-1.39708
S&P500	0.54333	1.86246	0.12423	0.65327	0.01159	0.02196	0.31844	0.02625

Tableau A.2: Résultats du maximum de vraisemblance pour le modèle MS-GARCH selon les indices composés du S&P/TSX et S&P500.

S&P/TSX	BS	GARCH(1,1)	RSLN(2 états)	MJD	MS-GARCH (2 états)
$\log f(y_{1:T}, \theta)$	4342.66	4585.21	4596.22	4512.85	4633.57
AIC	-8681.3	-9162.4	-9180.4	-9015.7	-9251.1
BIC	-8670.3	9140.3	-9147.4	-8988.2	-9207.1
S&P500	BS	GARCH(1,1)	RSLN(2 états)	MJD	MS-GARCH (2 états)
$\log f(y_{1:T}, \theta)$	4300.21	4498.11	4489.88	4415.42	4515.52
AIC	-8596.4	-8988.2	-8967.8	-8820.9	-9015.0
BIC	-8585.3	-8966.1	-8934.7	-8793.3	-8971.0

Tableau A.3: Résultats du maximum de vraisemblance pour les cinq modèles stochastiques selon les indices composés du S&P/TSX et S&P500.

RÉFÉRENCES

- Ait-Sahalia, Y. (2001). Closed-form likelihood expansions for multivariate diffusions. Working Paper, *Princeton University*.
- Ait-Sahalia, Y. & Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Finance Economics*, **88** : 413-452.
- American Academy of Actuaries (2005). Recommended approach for setting regulatory risk-based capital requirements for variable annuities and similar products. *National Association of Insurance Commissioners*.
- Arulampalam, S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, **50** : 174-189.
- Atiya, A. F. & Wall, S. (2009). An analytic approximation of the likelihood function for the Heston model volatility estimation problem. *Quantitative Finance*, **9** : 289-296.
- Augustyniak, M., Boudreault, M. & Morales, M. (2015). Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. *envoyé pour publication*.
- Bakshi, G., Cao, C. & Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, **52** : 205-238.
- Bams, D., Lehnert, T. & Wolff, C.C.P. (2009). Loss function in option valuation : *A framework for selection*. *Management Science*, **55** : 853-862.
- Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **3** : 637-654.
- Black, F. (1976). Studies of stoch price volatility changes. In : Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, *American Statistical Association*, 177-181.
- Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31** : 307-327.

- Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, **7** : 265-296.
- Broadie, M. & Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, **54** : 217-231.
- Canadian Institute of Actuaries, (2012). Calibration of equity returns for segregated fund liabilities. *Committee on Life Insurance Financial Reporting*.
- Clements, A.E., Hurn, A.S. & White, S.I. (2004a). Discretised non-linear filtering of dynamic latent variable models with application to stochastic volatility. Discussion paper no 179, *School of Economics, Finance and International Competitiveness*, Queensland University of technology.
- Clements, A.E. & White, S.I. (2004b). Non-linear filtering for stochastic volatility models with heavy tails and leverage. Discussion paper no 192, *School of Economics, Finance and International Competitiveness*, Queensland University of technology.
- Chernov, M. & Ghysels, M. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, **56** : 267-284.
- Chopin, N. (2004). Central limit theorem for sequential Monte Carlo and its application to Bayesian inference. *The Annals of Statistics*, **32** : 2385-2411.
- Christie, A. A. (1982). The stochastic behavior of common stock variances : Value, leverage and interest rate effects. *Journal of Financial Economics*, **10** : 407-432.
- Christoffersen, P. & Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, **72** : 291-318.
- Christoffersen, P., Heston, S. & Jacobs, K. (2009). The shape and term structure of the index option smirk : why multifactor stochastic volatility models work so well. *Management Science*, **55** : 1914-1932.
- Christoffersen, P., Jacobs K., Mimouni, K. (2010). Models for S&P500 volatility dynamics : evidence from realized volatility, daily return and option prices. *Review of Financial Studies*, **23** : 2139-2183.
- Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Review*, **31** : 245-296.
- Danielsson, J. & Richard, J.F. (1993). Accelerated Gaussian importance sampler

- with application to dynamic latent variable models. *Journal of Applied Econometrics*, **8** : 153–173.
- Danielsson, J. (1994). Stochastic volatility in asset prices : estimation with simulated maximum likelihood. *Journal of Econometrics*, **64** : 375-400.
- Del Moral, P. (1996). Nonlinear filtering : interacting particle solution. *Markov Processes and Related Fields*, **2** : 339-350.
- Doucet, A., de Freitas, J. F. G. & Gordon, N. (2001). Sequential Monte Carlo in Practice. *Springer-Verlag*, New York.
- Doucet, A. & Johansen, A. (2008). A tutorial on particle filtering and smoothing : fifteen years later. In : *The Oxford Handbook of Nonlinear Filtering*. Oxford. UK : Oxford University Press.
- Duffie, D. & Singleton, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, **61** : 929–952.
- Engle, R.F. and V. Ng (1993). Measuring and testing the impact of news on volatility *Journal of Finance*, **48** : 1749-1778.
- Eraker, B. (2004). Do stock prices and volatility jump ? Reconciling evidence from spot and option prices. *Journal of Finance*, **59** : 1367-1403.
- Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, **54** : 173-182.
- Fridman, M. & Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models, *Journal of Business and Economic Statistics*, **16** : 284–291.
- Geweke, J. (1994). Bayesian comparison of econometric models. Working Paper, *Federal Reserve Bank of Minneapolis*, Minnesota.
- Glasserman, P. (2003). Monte Carlo methods in financial engineering. New York, NY : *Springer*.
- Glosten, L. R., Jagannathan, R. Runkle, D. E., (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48** : 1770-1801.
- Gordon, N. J., Salmond, D. J. & Smith, A. F. M. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings, Part F : Radar and Sonar Navigation*, **140** : 107-113.
- Gray, Stephen F. (1996). Modeling the Conditional Distribution of Interest Rates

- as a Regime-Switching Process. *Journal of Financial Economics*, **42** : 27-62.
- Hardy, M. R., Freeland, R. K. & Till, M. C. (2006). Validation of long-term equity return models for equity-linked guarantees. *North American Actuarial Journal*, **10** : 28-47.
- Harvey, A. C., Ruiz, E. & Shephard, N. G. (1993). Multivariate stochastic variance models. *Review of Economic Studies*, **61** : 247-264.
- Harvey, A. C. & Shephard, N. G. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, **14** : 429-434.
- Hamilton, J.D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, **57** : 357-384.
- Hamilton, J. & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, **64** : 307-333.
- Heston, S. (1993). A closed-form solution for options with stochastic volatility with application to bonds and currency options. *Review of Financial Studies*, **6** : 327-343.
- Hüzeler, M. & Künsch, H. R. (2001). "Approximating and Maximising the Likelihood for a General State-Space Model" in Sequential Monte Carlo Methods in Practice, A. Doucet, J. F. G. de Freitas & N.J. Gordon *New York : Springer - Verlag*, **2001** : 159-175.
- Jacquier, E., Polson, N. & Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics*, **12** : 371-417.
- Jacquier, E., Polson, N. & Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, **122** : 185-212.
- Javaheri, A. (2005). Inside volatility arbitrage. *John Wiley & Sons*, New Jersey.
- Jones, C. S. (2003). The dynamics of stochastic volatility : Evidence from underlying and options markets. *Journal of Econometrics*, **116** : 181-224.
- Kim, S., Shephard, N. & Chib, S. (1994). Stochastic volatility. Optimal Likelihood Inference and Comparison with ARCH Models. (Unpublished Paper, *Nuffield College, Oxford*).
- Kim, S. & Shephard, N. (1998). Stochastic volatility : likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **45** : 361-393.

- Kitawaga, G., 1987, Non-gaussian state-space modeling of non-stationary time series. *Journal of the American Statistical Association*, **82** : 1032-1062.
- Knight, J. L., Satchell, S. E. & Yu, J. (2002). Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian and New Zealand Journal of Statistics*, **44** : 319-335.
- Lewis, A. L. (2000). Option valuation under stochastic volatility : with Mathematica code. *Finance Press*.
- MacDonald, I. & Zucchini, W. (1997). Hidden markov and other models for discrete-valued time series. *Chapman and Hall*, London.
- Malik, S. & Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, **165** : 190-209.
- Melino, A. & Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, **45** : 239-265.
- Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **3** : 125-44.
- Nelson, D. B. (1988). The time series behavior of stock market volatility and returns. (Unpublished Ph D Dissertation, *Massachusetts Institute of Technology*, Cambridge, MA).
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns : A New Approach. *Econometrica*, **59** : 347-370.
- Pitt, M. K. (2002). Smooth particle filters for likelihood evaluation and maximisation. Unpublished manuscript, *Departement of Economics*, Warwick University.
- Pitt, M. K., Malik, S. & Doucet, A. (2014). Simulated likelihood inference for stochastic volatility models using continuous particle filtering. *Journal of Econometrics*, **66** : 527-552.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, **63** : 289-306.
- Sandmann, G. & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, **87** : 271-301.
- Shephard, N. G. (1993). Fitting non-linear time series models, with application to stochastic variance models. *Journal of Applied Econometrics*, **8** : 135-152.
- Singleton, K. J. (2001). Estimation of affine asset pricing models using the empi-

- rical characteristic function. *Journal of Econometrics*, **102** : 111–141.
- Smith, D. R. (2009). Asymmetry in Stochastic Volatility Models : Threshold or Correlation?. *Studies in Nonlinear Dynamics & Econometrics*, **13** : Article 1.
- So, M. K., Li, W. & Lam, K. (2003). On a threshold stochastic volatility model. *Journal of Forecasting*, **22** : 473-500.
- Taylor, S. (1986). Modeling financial time series. *Chichester : Wiley*.
- Tsay, R. (2001). Analysis of Financial Time Series. *John Wiley & Sons*, 1st edition.
- Van Haastrecht, A. & Pelsser, A. (2010). Efficient, almost exact simulation of the Heston stochastic volatility model. *International Journal of Theoretical and Applied Finance*, **13** : 1-43.
- Watanabe, T. (1999). A non-linear filtering approach to stochastic volatility models with an application to daily stock returns. *Journal of Applied Econometrics*, **14** : 101–121.