

Université de Montréal

**Estimation des modèles à volatilité stochastique par
l'entremise du modèle à chaîne de Markov cachée**

par

Jean Hounkpe

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en statistique

11 janvier 2018

SOMMAIRE

La problématique d'estimation des paramètres des modèles à volatilité stochastique par maximisation directe de la vraisemblance est adressée. À cet effet, nous présentons un algorithme approximant numériquement le filtre optimal à partir de la méthodologie proposée par [Kitagawa \(1987\)](#) pour la résolution du problème de filtrage dans les systèmes non-linéaires et/ou non-gaussiens.

Nous montrons que cet algorithme consiste à exécuter le filtre d'Hamilton (le filtre d'Hamilton offre une solution optimale au problème de filtrage pour un modèle à chaîne de Markov cachée) sur une discrétisation de la variable latente continue. La solution proposée améliore considérablement le temps de calcul et produit des résultats au moins aussi bons que les approches de référence dans le domaine.

Par la suite, nous présentons et démontrons une généralisation de cet algorithme au cas des modèles à volatilité stochastique incorporant un effet de levier et des sauts. Plusieurs études Monte Carlo et empiriques sont réalisées pour évaluer la qualité de l'approche dans l'approximation de la log-vraisemblance et l'estimation des paramètres des différents modèles à volatilité stochastique. Nous présentons aussi une comparaison de cette approche à celle par *filtre particulière continue*.

Mots-clés : Volatilité stochastique, Modèle à chaîne de Markov cachée, Filtrage non-linéaire et non-gaussien, Intégration numérique, Maximum de vraisemblance, Filtre particuliers, Effet de levier, Sauts.

SUMMARY

The problem of estimating the parameters of stochastic volatility models by direct maximisation of the likelihood is addressed. To this end, we present an algorithm that numerically approximates the optimal filter from the methodology proposed by [Kitagawa \(1987\)](#) for solving the filtering problem in non-linear and/or non-Gaussian systems.

We show that this algorithm corresponds to running the Hamilton filter (the Hamilton filter offers an optimal solution to the filtering problem for a hidden Markov model on a finite state space) on a discretization of the continuous latent variable. The proposed solution significantly improves the computation time and produces results at least as good as state-of-the-art approaches in the field.

Subsequently, we present and demonstrate a generalization of this algorithm in the case of stochastic volatility models incorporating leverage and jumps. Several Monte Carlo and empirical studies are conducted to evaluate the quality of the approach for approximating the log-likelihood and estimating the parameters. We also present a comparison of this approach to the *continuous particle filter* approach.

Keywords : Stochastic volatility, Hidden Markov model, Non-linear and non-Gaussian filtering, Numerical integration, Maximum likelihood, Particle filter, Leverage effect, Jumps.

TABLE DES MATIÈRES

Sommaire	iii
Summary	v
Liste des figures	xi
Liste des tableaux	xiii
Remerciements	xv
Introduction	1
Chapitre 1. Modèles à volatilité stochastique : introduction et méthodes d'estimation	7
1.1. Modèle à volatilité stochastique standard : spécification et propriétés.....	8
1.1.1. Modèle SV_0	8
1.1.2. Propriétés stationnaires du modèle SV_0	9
1.2. Extensions du modèle à volatilité stochastique standard.....	13
1.2.1. Modèle SVL.....	15
1.2.2. Modèle SVLJ.....	16
1.3. Estimation des modèles à volatilité stochastique.....	17
1.3.1. Méthode des moments généralisées (GMM).....	18
1.3.2. Quasi-maximum de vraisemblance (QML).....	19
1.3.3. Méthodes bayésiennes.....	22
1.3.4. Méthodes du maximum de vraisemblance.....	22
Chapitre 2. Approche d'estimation des modèles à volatilité stochastique basée sur le modèle à chaîne de Markov cachée	27
2.1. Modèles espace-état : problème de filtrage.....	27
2.1.1. Problème de filtrage.....	28
2.1.2. Filtre optimal.....	29

2.2.	Évaluations exactes du filtre optimal.....	30
2.2.1.	Filtre de Kalman.....	30
2.2.2.	Filtre d'Hamilton ou en espace d'états discret (<i>grid-based filter</i>).....	32
2.3.	Approximations du filtre optimal.....	35
2.3.1.	Filtre par approximation numérique de la distribution filtrée (<i>Filtre HMM</i>).....	35
2.3.2.	Filtre particulaire.....	39
2.4.	Approximations de la vraisemblance des modèles à volatilité stochastique.....	46
2.4.1.	Estimation des modèles à volatilité stochastique par le <i>filtre HMM</i>	46
2.4.2.	Estimation des modèles à volatilité stochastique par <i>filtre particulaire continu</i>	52
Chapitre 3.	Études par simulations de la qualité de l'approximation de la vraisemblance	55
3.1.	Considérations numériques d'implémentation.....	55
3.2.	Qualité de l'approximation de la log-vraisemblance du modèle <i>AR(1)</i> bruité.....	57
3.2.1.	Étude Monte Carlo sur le modèle <i>AR(1)</i> bruité.....	57
3.2.2.	Résultats des études sur l'approximation de la log-vraisemblance du modèle <i>AR(1)</i> bruité.....	58
3.3.	Comparaison des approximations de log-vraisemblance des modèles à volatilité stochastique.....	63
3.3.1.	Données et jeux de paramètres.....	63
3.3.2.	Résultats des études sur l'approximation de la log-vraisemblance des modèles <i>SV₀</i> , <i>SVL</i> et <i>SVLJ</i>	64
Chapitre 4.	Évaluation de la qualité de l'estimation des paramètres des modèles à volatilités stochastiques	69
4.1.	Performances des deux approches dans l'estimation des paramètres du modèle <i>SV₀</i>	69
4.1.1.	Expérience de simulation JPR.....	70
4.1.2.	Présentation des résultats de l'étude JPR.....	71
4.2.	Performance des deux approches dans l'estimation des paramètres des modèles <i>SVL</i> et <i>SVLJ</i>	71

4.2.1. Expérience de simulation	73
4.2.2. Présentation des résultats de l'expérience	73
Chapitre 5. Estimations sur des données empiriques	77
5.1. Estimation des paramètres	77
5.2. Qualité d'adéquation aux données	81
Conclusion	83
Bibliographie	87
Annexe A. Preuve de la proposition 2.2.1	A-i
Annexe B. Introduction aux modèles à chaîne de Markov cachée	B-i
B.1. Chaîne de Markov à espace d'états fini	B-i
B.2. Modèles à chaîne de Markov cachée	B-iii

LISTE DES FIGURES

1.1	Rendements sur le S&P 500 au cours de la période allant du 2000-01-03 au 2016-12-30. <i>Haut-gauche</i> : évolution du rendement ; <i>Haut-droit</i> : normale Q-Q plot des rendements ; <i>Bas-gauche</i> : $\text{Corr}[y_{t+h}, y_t]$; <i>Bas-droit</i> : $\text{Corr}[y_{t+h} , \max(-y_t, 0)]$ en noir ; <i>Bas-droit</i> : $\text{Corr}[y_{t+h} , \max(y_t, 0)]$ en rouge.	14
3.1	Box-plot des erreurs d'approximations de la log-vraisemblance du modèle $AR(1)$ bruité par <i>filtre HMM</i> , selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k).....	61

LISTE DES TABLEAUX

3.1	Qualité d'approximation de la log-vraisemblance d'un modèle $AR(1)$ bruité par <i>filtre HMM</i> , selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k) (Début).	59
3.2	Qualité d'approximation de la log-vraisemblance d'un modèle $AR(1)$ bruité par <i>filtre HMM</i> , selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k) (Fin).	60
3.3	Qualité d'approximation de la log-vraisemblance du modèle $AR(1)$ bruité par <i>filtre particulaire continu (CPF)</i> et <i>bootstrap filter (PF)</i> selon le nombre de particules (M).	62
3.4	Comparaison de l'approximation de la log-vraisemblance du modèle SV_0 , selon les approches par filtres HMM et par filtres particulaires, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM_1 , HMM_2 ou HMM_3). Le paramètre k est fixé à 5.	66
3.5	Comparaison de l'approximation de la log-vraisemblance du modèle SVL, selon les approches par filtres HMM et par filtres particulaires, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM_1 , HMM_2 ou HMM_3). Le paramètre k est fixé à 5.	67
3.6	Comparaison de l'approximation de la log-vraisemblance du modèle SVLJ, selon les approches par filtres HMM et par filtres particulaires, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM_1 , HMM_2 ou HMM_3). Le paramètre k est fixé à 5.	68
4.1	Comparaison des approches MCMC, MCL, F&H, <i>filtre particulaire continu (CPF)</i> avec $M = 500$ particules et <i>filtre HMM</i> (HMM_1) avec $N = 30$ intervalles pour l'estimation des paramètres du modèle SV_0	72
4.2	Résultats de l'estimation des paramètres du modèle SVL par <i>filtre particulaire continu (CPF)</i> avec $M = 500$ particules et <i>filtre HMM</i> (HMM_1) avec $N = 30$ intervalles sur les données simulées pour deux jeux de paramètres.	74

4.3	Résultats de l'estimation des paramètres du modèle SVLJ par <i>filtre particulaire continu</i> (CPF) avec $M = 500$ particules et <i>filtre HMM</i> (HMM_1) avec $N = 30$ intervalles sur les données simulées pour trois différents jeux de paramètres..	75
5.1	Résultats de l'estimation des paramètres des modèles à volatilité stochastique par <i>filtre particulaire continu</i> (CPF) et <i>filtre HMM</i> sur la série S&P 500 pour la période du 2000-01-03 au 2016-06-30.....	78
5.2	Résultats de l'estimation des paramètres des modèles à volatilité stochastique par <i>filtre particulaire continu</i> (CPF) et <i>filtre HMM</i> sur la série NASDAQ pour la période du 2000-01-03 au 2016-06-30.....	79
5.3	Résultats de l'estimation des paramètres des modèles à volatilité stochastique par <i>filtre particulaire continu</i> (CPF) et <i>filtre HMM</i> (HMM_1) sur la série USD/EUR pour la période du 2000-01-03 au 2016-06-30.....	80
5.4	Comparaison de la qualité d'adéquation de différentes modélisations avec effet de levier de la volatilité des trois séries financières.....	82

REMERCIEMENTS

Je voudrais avant toutes choses formuler mes sincères remerciements et toute ma gratitude à mon superviseur, professeur Maciej Augustyniak, pour son implication dans cette recherche, sa confiance, ses encouragements et conseils et pour son soutien financier.

Je suis également reconnaissant au Département de mathématiques et de statistique et en particulier au professeur Pierre Duchesne pour son soutien financier.

J'aimerais terminer ces remerciements pour réitérer ma profonde gratitude à mes parents, mes frères et sœurs, à ma compagne et à la famille Fanton pour leur soutien tout au long de mes études.

INTRODUCTION

La qualité d'un modèle réside dans sa simplicité et sa capacité à générer les caractéristiques de la dynamique du phénomène étudié. La reproduction des nombreux faits stylisés (voir : [Cont \(2001\)](#) pour une présentation détaillée des faits stylisés en finance) présents dans la plupart des séries financières par des modèles économétriques est l'un des grands défis dans le domaine de la finance. Depuis le début des années 1960, les travaux de [Osborne \(1959\)](#), [Mandelbrot \(1963, 1971\)](#), [Fama \(1965\)](#), [Engle \(1982\)](#), [Rubinstein \(1985\)](#) et [Bollerslev et al. \(1992\)](#) ont révélé, d'une part, que les rendements de faible et de forte amplitudes tendent à être regroupés (*clustering de volatilité* et *hétéroscédasticité conditionnelle*) et, d'autre part, que la distribution des rendements des actifs financiers est plus épaisse aux extrémités et plus pointue au centre comparativement à une distribution normale (*leptokurtique*).

Ces caractéristiques des séries financières remettent en question l'hypothèse de constance de la volatilité conditionnelle des rendements stipulée dans bon nombre de théories financières à l'instar de la théorie du portefeuille de [Markowitz \(1952\)](#) et celle du modèle de Black & Scholes ([Black et Scholes \(1973\)](#)). Compte tenu de son rôle capital en finance (voir : [Tsay \(2010\)](#), page 106), des efforts considérables ont dès lors été consacrés par les chercheurs sur la modélisation de la volatilité des rendements des actifs financiers.

Deux différentes approches existent dans la littérature pour modéliser l'évolution dans le temps de la volatilité des rendements d'un actif financier. La première consiste à considérer que cette volatilité est déterministe et s'écrit comme une fonction des erreurs et des variances

passées. C'est la classe des modèles *autoregressive conditional heteroscedasticity* (ARCH) introduite en 1982 par Engle (1982) et étendue en 1986 par Bollerslev (1986). La famille des modèles à volatilité stochastique constitue la seconde approche, dans laquelle un terme d'erreur aléatoire additionnel est introduit dans la dynamique de la variance conditionnelle, qui devient ainsi un processus stochastique latent. Taylor (1982, 1986) fut le premier à proposer ce type de modélisation comme alternative aux modèles ARCH. Ghysels et al. (1996) et Shephard (1996) peuvent être consultés pour une revue sur les modèles à volatilité stochastique et Bollerslev (2010) et Bollerslev et al. (1994) en ce qui concerne les modèles *generalized autoregressive conditional heteroscedasticity* (GARCH).

Cette innovation additionnelle, introduite dans la dynamique des modèles à volatilité stochastique, rend ces derniers non seulement plus flexibles, mais aussi plus adaptés aux données empiriques comparativement aux modèles GARCH (Melino et Turnbull (1990); Carnero et al. (2004)). La supériorité des modèles à volatilité stochastique sur les modèles GARCH a été relevée dans plusieurs travaux comme Danielsson (1994), Kim et al. (1998) et Carnero et al. (2004). Toutefois, le modèle GARCH ainsi que ses nombreuses extensions sont généralement préférés aux modèles à volatilité stochastique à cause des difficultés engendrées par l'inobservation de cette volatilité stochastique sur l'évaluation de la vraisemblance du modèle. En effet, la vraisemblance du modèle à volatilité stochastique est une intégrale multidimensionnelle puisqu'elle nécessite l'intégration par rapport aux valeurs possibles de cette variable latente qu'est la volatilité. Toute la difficulté de l'estimation des modèles à volatilité stochastique par l'approche classique de maximisation de la vraisemblance provient du fait qu'il est impossible de calculer analytiquement cette intégrale multidimensionnelle de même dimension que la taille des données (voir : Jacquier et al. (1994), page 371).

Face à cette difficulté, plusieurs outils économétriques ont été développés, dès la fin des années 1980, pour faciliter l'estimation des paramètres de cette classe de modèles. Ces

outils peuvent être regroupés en deux grandes catégories : celle des approches basées sur les moments du modèle et celle des approches basées sur la vraisemblance du modèle.

Dans la première catégorie on retrouve les approches exploitant la technique de la méthode des moments. L'avantage principal de cette méthodologie dans le cadre des modèles à volatilité stochastique est de permettre l'estimation des paramètres sans passer par l'évaluation de cette intégrale multidimensionnelle. [Taylor \(1986\)](#) a été le premier à proposer cette approche pour l'estimation des paramètres dans le cadre de ce type de modèle. À l'instar de [Jacquier et al. \(1994\)](#), plusieurs auteurs ont montré que l'estimation des modèles à volatilité stochastique par la méthode des moments produit des estimateurs biaisés (voir : [Andersen \(1994\)](#), [Ghysels et Jasiak \(1994\)](#)).

La seconde catégorie regroupe, d'une part, les méthodes bayésiennes, et, d'autre part, les méthodes du maximum de vraisemblance. L'approche bayésienne vise à estimer le vecteur des paramètres du modèle à travers la simulation de sa distribution à posteriori. Basée sur les méthodes Monte-Carlo par chaînes de Markov (MCMC), l'approche bayésienne a été introduite par [Jacquier et al. \(1994\)](#) pour l'estimation des modèles à volatilité stochastique. L'idée principale des MCMC est de construire récursivement une chaîne de Markov, ayant comme distribution stationnaire la distribution à posteriori, par le biais d'une méthode d'échantillonnage appropriée. La performance et la rapidité de l'approche dépendent fortement de la méthode d'échantillonnage adoptée. Comme l'ont si bien souligné Shephard (voir : [Shephard \(2005\)](#), page 13) et Fridman et Harris (voir : [Fridman et Harris \(1998\)](#), page 285), les deux plus grands désavantages de ces méthodes se trouvent dans leur exigence en ressources computationnelles et en temps de calcul ainsi que dans la complexité de les implémenter. Les approches bayésiennes ont largement été utilisées dans la littérature des modèles à volatilité stochastique ; les articles de [Chib et al. \(2002\)](#) et de [Jacquier et al. \(2004\)](#) peuvent être consultés sur le sujet.

Comparativement aux autres approches, les méthodes du maximum de vraisemblance ont été peu exploitées pour fin d'estimation des paramètres des modèles à volatilité stochastique à cause de l'impossibilité d'évaluer analytiquement la vraisemblance de ce type de modèle. Deux types de solution peuvent être utilisées à cette fin : stochastique ou déterministe. L'approche stochastique consiste en l'utilisation de la technique de simulation Monte Carlo séquentielle (*sequential Monte Carlo*) ou filtre particulaire. L'approximation de la vraisemblance obtenue par cette méthode n'est cependant pas une fonction continue des paramètres, ce qui rend inefficace leur estimation par sa maximisation. Une version continue du filtre particulaire a récemment été proposée par [Malik et Pitt \(2011\)](#) et utilisée par [Pitt et al. \(2014\)](#) pour l'estimation des modèles à volatilité stochastique. Le désavantage majeur de cette méthode non-déterministe est le coût en temps de calcul.

L'approche déterministe consiste en l'utilisation de procédures d'intégration numérique récursive inspirées de la méthodologie introduite par [Kitagawa \(1987\)](#). [Fridman et Harris \(1998\)](#) ont adopté cette méthode d'approximation dans le cadre de l'estimation des paramètres des modèles à volatilité stochastique. La complexité et la difficulté d'implémentation des méthodes d'intégration déterministe ont relativement limité l'adoption de cette approche dans ce domaine. Le lien entre les modèles à chaîne de Markov cachée et cette approche déterministe a été récemment souligné par [Langrock et al. \(2012\)](#) ce qui permet de faciliter son implémentation. La qualité de cette approximation pour le calcul de la log-vraisemblance ainsi que pour l'estimation des paramètres reste toutefois à étudier.

Notre objectif dans le présent mémoire est de contribuer à cette littérature à travers :

- la dérivation et la formulation de cette approximation par l'entremise des modèles à chaîne de Markov cachée,
- l'étude approfondie de la qualité d'approximation de cette méthode déterministe pour le calcul de la log-vraisemblance du modèle à volatilité stochastique standard,
- l'extension aux modèles à volatilité stochastique avec effet de levier et sauts,

- l'étude de la qualité des estimateurs et comparaison avec l'approche du filtre particulière continu (Pitt et al. (2014)).

Dans le premier chapitre, nous allons introduire le modèle à volatilité stochastique standard ainsi que deux extensions permettant d'incorporer plus de faits stylisés. On y présente aussi les principales approches d'estimation des paramètres de ces modèles. Le chapitre 2 est consacré à la présentation des différentes approches pour approximer la vraisemblance des modèles à volatilité stochastique. Nous avons dans un premier temps introduit le cadre général des modèles espace-état et le problème de filtrage. Les solutions exactes (filtre de Kalman, filtre d'Hamilton) et approchées (*filtre HMM (hidden Markov model)*, *filtre particulière continu*) de résolution de ce problème ont été par la suite présentées. Nous concluons ce second chapitre en présentant l'algorithme d'approximation de la vraisemblance des modèles à volatilité stochastique.

Les trois derniers chapitres ont été consacrés à l'évaluation de la qualité et la comparaison des approches par *filtre HMM* et par *filtre particulière continu* dans l'approximation de la log-vraisemblance et l'estimation des paramètres des modèles à volatilité stochastique sur des données simulées et empiriques. Ainsi, le chapitre 3 est consacré à l'étude de la qualité de l'approximation de la log-vraisemblance des modèles espace-état par les approches du *filtre HMM* et du *filtre particulière*. Nous avons considéré dans un premier temps l'approximation de la log-vraisemblance du modèle $AR(1)$ gaussien bruité. Par la suite, nous avons comparé les approximations de la log-vraisemblance des modèles à volatilité stochastique sur la série du *S&P 500* par ces deux filtres. L'évaluation de la performance de ces deux approches pour l'estimation par maximum de vraisemblance des paramètres des modèles à volatilité stochastique fait l'objet du chapitre 4. Dans un premier temps, une réplique de l'étude de [Jacquier et al. \(1994\)](#) a été menée afin d'évaluer et de comparer les résultats des approches par *filtre HMM* et par *filtre particulière continu* aux méthodes MCMC de [Jacquier et al. \(1994\)](#), de maximum de vraisemblance de [Fridman et Harris \(1998\)](#) et de *Monte Carlo*

Likelihood (MCL) de [Sandmann et Koopman \(1998\)](#). Par la suite, nous avons réalisé un second groupe d'études à travers la reproduction des tableaux 2, 3 et 4 de l'article de [Pitt et al. \(2014\)](#). Et enfin, dans le dernier chapitre nous avons présenté une comparaison de la qualité d'adéquation aux données des modèles à volatilité stochastique par rapport aux modèles GARCH et MS-GARCH (GARCH avec changement de régimes) sur les données de rendements des indices S&P 500 et NASDAQ et du taux de change USD/EUR.

Chapitre 1

MODÈLES À VOLATILITÉ STOCHASTIQUE : INTRODUCTION ET MÉTHODES D'ESTIMATION

La volatilité du rendement des actifs financiers n'étant pas directement observable sur les marchés, il est capital de développer des modèles permettant d'évaluer avec le plus de précision possible cette quantité. Dans cette quête, les modèles à volatilité stochastique sont des candidats de choix, puisqu'en plus des raisons annoncées précédemment comparativement aux modèles GARCH, ces modèles s'intègrent aisément aux fondements théoriques de la finance moderne sur l'évaluation des options ([Taylor \(1994\)](#)), d'une part, et d'autre part, dans la famille des modèles espace-état (voir : [Durbin et Koopman \(2012\)](#), page 216).

Dans ce chapitre, nous allons dans un premier temps introduire le modèle à volatilité stochastique standard et présenter deux extensions naturelles permettant d'incorporer plus de faits stylisés. La présentation des principales approches d'estimation des paramètres de ce modèle fera l'objet de la seconde partie.

1.1. MODÈLE À VOLATILITÉ STOCHASTIQUE STANDARD : SPÉCIFICATION ET PROPRIÉTÉS

1.1.1. Modèle SV_0

Dans sa forme la plus simple, telle qu'introduite par [Taylor \(1986\)](#), le modèle à volatilité stochastique standard (SV_0) est spécifié pour une série $y_{1:T} = (y_1, y_2, \dots, y_T)'$ de rendements observés aux temps t ($t = 1, \dots, T$) comme suit :

$$y_t = \exp\left(\frac{1}{2}h_t\right) \epsilon_t, \quad (1.1.1a)$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \eta_t, \quad (1.1.1b)$$

$$\text{avec } \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

où $\{h_t\}$ est un processus latent (non observé) et représente le logarithme de la volatilité du rendement au temps t , $\{\epsilon_t\}$ et $\{\eta_t\}$ sont deux suites de processus indépendants de distributions normales centrées en zéro et de variances égales à l'unité ($\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ et $\eta_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$). La corrélation entre les variables ϵ_t et η_t est mesurée par le paramètre ρ . Il est possible de réécrire l'expression (1.1.1b) à l'aide d'une innovation ξ_t indépendante de l'innovation ϵ_t grâce à la décomposition de Cholesky comme suit :

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \rho \epsilon_t + \sigma_\eta \sqrt{1 - \rho^2} \xi_t, \quad t = 0, \dots, T. \quad (1.1.2)$$

Dans le cadre du modèle à volatilité stochastique standard les innovations ϵ_t et η_t sont supposées indépendantes, soit $\rho = 0$.

La spécification suivante est équivalente et parfois adoptée dans la littérature pour les expressions (1.1.1a) et (1.1.1b) :

$$y_t = \sigma_t \epsilon_t,$$

$$\log \sigma_{t+1}^2 = \alpha + \phi \log \sigma_t^2 + \sigma_\eta \eta_t.$$

Le niveau du paramètre σ_η permet de modéliser l'incertitude liée à la log-volatilité future. Les composantes du vecteur des paramètres $\boldsymbol{\theta} = (\alpha, \phi, \sigma_\eta)'$ avec $\alpha \in (-\infty, \infty)$, $\phi \in (-1, 1)$ et $\sigma_\eta \in (0, \infty)$ peuvent respectivement être interprétées comme des paramètres de position, de persistance et d'échelle de la log-volatilité. Un des avantages théoriques du modèle SV_0 (1.1.1) sur le modèle GARCH est qu'il peut être considéré comme une approximation en temps discret du modèle à volatilité stochastique continu proposé par [Hull et White \(1987\)](#) comme extension de la théorie de [Black et Scholes \(1973\)](#) pour l'évaluation des options.

1.1.2. Propriétés stationnaires du modèle SV_0

Sous l'hypothèse $\eta_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, la log-volatilité $\{h_t\}$ est un processus $AR(1)$ (autorégressif d'ordre 1) gaussien comme le montre la relation (1.1.1b). La covariance et la stricte stationnarité du processus $\{h_t\}$ sont donc assurées pour $|\phi| < 1$ et dans le cas échéant, la relation (1.1.1b) est complétée par h_0 tiré de la distribution stationnaire $h_0 \sim N(\mu_h, \sigma_h^2)$ avec :

$$\mathbb{E}[h_t] = \mu_h \implies \mu_h = \alpha + \phi \mu_h \text{ soit } \mu_h = \frac{\alpha}{1 - \phi} \quad (1.1.3)$$

$$\text{Var}[h_t] = \sigma_h^2 \implies \sigma_h^2 = \phi^2 \sigma_h^2 + \sigma_\eta^2 \text{ soit } \sigma_h^2 = \frac{\sigma_\eta^2}{1 - \phi^2}. \quad (1.1.4)$$

Proposition 1.1.1. *Lorsque le processus $\{h_t\}$, dans le cadre du modèle SV_0 , est covariance (resp. strictement) stationnaire, le processus $\{y_t\}$ est aussi covariance (resp. strictement) stationnaire et ses moments d'ordre impair ainsi que sa fonction d'auto-covariance sont nulles.*

DÉMONSTRATION. Comme produit indépendant du bruit blanc gaussien $\{\epsilon_t\}$ et du processus $\{\exp(h_t/2)\}$, le processus $\{y_t\}$ est stationnaire si et seulement si le processus $\{h_t\}$ est stationnaire. Ainsi, $\{\epsilon_t\}$ étant un processus de distribution symétrique par rapport à zéro

(bruit blanc), tous les moments d'ordre impair du processus $\{y_t\}$ sont nuls. Par indépendance entre $\{\epsilon_t\}$ et $\{h_t\}$, la fonction d'auto-covariance de $\{y_t\}$ est donnée par :

$$\text{Cov}[y_t, y_{t+\tau}] = \mathbb{E}[y_t y_{t+\tau}] = \mathbb{E}\left[\exp\left(\frac{1}{2}(h_t + h_{t+\tau})\right)\right] \mathbb{E}[\epsilon_t \epsilon_{t+\tau}] = 0, \quad \tau = 1, 2, \dots \quad (1.1.5)$$

□

Proposition 1.1.2. *Lorsque le processus $\{h_t\}$ est stationnaire, tous les moments d'ordre pair ($r = 2, 4, \dots$) du processus $\{y_t\}$ existent et sont donnés par :*

$$\mathbb{E}[y_t^r] = \mathbb{E}\left[\exp(rh_t/2)\right] \mathbb{E}[\epsilon_t^r] = \exp\left(\frac{r}{2}\mu_h + \frac{r^2}{8}\sigma_h^2\right) \frac{r!}{2^{\frac{r}{2}}(r/2)!}. \quad (1.1.6)$$

DÉMONSTRATION. Le processus $v_t := \exp(h_t)$ étant de distribution log-normale, nous avons pour $k \in \mathbb{N}$: $\mathbb{E}[v_t^k] = \mathbb{E}\left[\exp(kh_t)\right] = \exp\left(k\mu_h + \frac{k^2}{2}\sigma_h^2\right)$. Aussi, pour $\epsilon_t \sim N(0,1)$ nous avons : $\mathbb{E}[\epsilon_t^{2k}] = \frac{(2k)!}{2^k k!}$. En effet, par intégration par partie on a :

$$\begin{aligned} \mathbb{E}[\epsilon_t^{2(k+1)}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \epsilon_t^{(2k+1)} \epsilon_t e^{-\frac{\epsilon_t^2}{2}} d\epsilon_t \\ &= \lim_{x \rightarrow \infty} \left(\frac{1}{\sqrt{2\pi}} \left[-\epsilon_t^{(2k+1)} e^{-\frac{\epsilon_t^2}{2}} \right]_{-x}^x \right) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (2k+1) \epsilon_t^{2k} e^{-\frac{\epsilon_t^2}{2}} d\epsilon_t \\ &= (2k+1) \mathbb{E}[\epsilon_t^{2k}]. \end{aligned} \quad (1.1.7)$$

Pour $k = 1$, on a $\mathbb{E}[\epsilon_t^2] = 1 = \frac{2!}{2^1(1)!}$ et en supposant la relation $\mathbb{E}[\epsilon_t^{2k}] = \frac{(2k)!}{2^k k!}$ vraie à l'ordre $k = 2, 3, \dots$, on déduit de (1.1.7) :

$$\begin{aligned} \mathbb{E}[\epsilon_t^{2(k+1)}] &= (2k+1) \mathbb{E}[\epsilon_t^{2k}] \\ &= (2k+1) \frac{(2k)!}{2^k k!} \\ &= \frac{(2k+2)(2k+1)(2k)!}{2(k+1)2^k k!} \end{aligned}$$

$$= \frac{(2(k+1))!}{2^{k+1}(k+1)!}, \quad (1.1.8)$$

ce qui nous permet de conclure par induction que $\mathbb{E}[\epsilon_t^{2k}] = \frac{(2k)!}{2^k k!}$ pour tout $k \in \mathbb{N}$. Il vient alors que tous les moments d'ordre pair ($r = 2, 4, \dots$) de $\{y_t\}$ existent lorsque $\{h_t\}$ est stationnaire et sont donnés par :

$$\mathbb{E}[y_t^r] = \mathbb{E}\left[\exp(rh_t/2)\right]\mathbb{E}[\epsilon^r] = \exp\left(\frac{r}{2}\mu_h + \frac{r^2}{8}\sigma_h^2\right)\frac{r!}{2^{\frac{r}{2}}(r/2)!}.$$

□

En particulier, les moments d'ordre deux et quatre de $\{y_t\}$ sont donnés par :

$$\mathbb{E}[y_t^2] = \exp(\mu_h + \frac{1}{2}\sigma_h^2), \quad (1.1.9)$$

$$\mathbb{E}[y_t^4] = 3\exp(2\mu_h + 2\sigma_h^2). \quad (1.1.10)$$

Proposition 1.1.3. *La variance $\text{Var}[y_t^2]$ et la covariance $\text{Cov}[y_t^2, y_{t+\tau}^2]$ du processus des rendements au carré $\{y_t^2\}$ sont données par :*

$$\text{Var}[y_t^2] = \exp(2\mu_h + \sigma_h^2)\left(3\exp(\sigma_h^2) - 1\right), \quad (1.1.11)$$

$$\text{Cov}[y_t^2, y_{t+\tau}^2] = \exp(2\mu_h + \sigma_h^2)\left(\exp(\sigma_h^2\phi^\tau) - 1\right), \tau = 1, 2, \dots \quad (1.1.12)$$

DÉMONSTRATION. Nous avons $y_t^2 = v_t\epsilon_t^2$, avec $v_t = \exp(h_t)$ et $\mathbb{E}[v_t^k] = \exp\left(k\mu_h + \frac{k^2}{2}\sigma_h^2\right)$ pour $k \in \mathbb{N}$. Par indépendance entre v_t et ϵ_t il vient alors :

$$\begin{aligned} \text{Var}[y_t^2] &= \mathbb{E}[y_t^4] - \mathbb{E}[y_t^2]^2 \\ &= \mathbb{E}[v_t^2\epsilon_t^4] - \mathbb{E}[v_t\epsilon_t^2]^2 \\ &= 3\mathbb{E}[v_t^2] - \mathbb{E}[v_t]^2 \\ &= 3\exp(2\mu_h + 2\sigma_h^2) - \exp(2\mu_h + \sigma_h^2) \end{aligned}$$

$$= \exp(2\mu_h + \sigma_h^2) \left(3 \exp(\sigma_h^2) - 1 \right).$$

Maintenant, en notant que :

$$\begin{aligned}
h_{t+\tau} &= \alpha + \phi h_{t+\tau-1} + \sigma_\eta \eta_{t+\tau-1} \\
&= \alpha + \phi(\alpha + \phi h_{t+\tau-2} + \sigma_\eta \eta_{t+\tau-2}) + \sigma_\eta \eta_{t+\tau-1} \\
&= \phi^2 h_{t+\tau-2} + \phi^1(\alpha + \sigma_\eta \eta_{t+\tau-2}) + \phi^0(\alpha + \sigma_\eta \eta_{t+\tau-1}) \\
&\vdots \\
&= \phi^\tau h_t + \sum_{i=0}^{\tau-1} (\alpha + \sigma_\eta \eta_{t-1-i}) \phi^i,
\end{aligned} \tag{1.1.13}$$

on peut écrire :

$$\begin{aligned}
\mathbb{E}[y_t^2 y_{t+\tau}^2] &= \mathbb{E}[v_t \epsilon_t^2 v_{t+\tau} \epsilon_{t+\tau}^2] \\
&= \mathbb{E}[v_t v_{t+\tau}] \\
&= \mathbb{E} \left[\exp(\phi^\tau h_t) \exp(h_t) \exp \left(\sum_{i=0}^{\tau-1} (\alpha + \sigma_\eta \eta_{t-1-i}) \phi^i \right) \right] \\
&= \mathbb{E}[v_t^{1+\phi^\tau}] \mathbb{E} \left[\exp \left(\sum_{i=0}^{\tau-1} (\alpha + \sigma_\eta \eta_{t-1-i}) \phi^i \right) \right] \\
&= \exp \left((1 + \phi^\tau) \mu_h + \frac{1}{2} (1 + \phi^\tau)^2 \sigma_h^2 \right) \exp \left(\sum_{i=0}^{\tau-1} \left(\alpha \phi^i + \frac{1}{2} \sigma_\eta^2 \phi^{2i} \right) \right) \\
&= \exp \left((1 + \phi^\tau) \mu_h + \frac{1}{2} (1 + \phi^\tau)^2 \sigma_h^2 + \frac{\alpha(1 - \phi^\tau)}{1 - \phi} + \frac{1}{2} \frac{\sigma_\eta^2 (1 - \phi^{2\tau})}{1 - \phi^2} \right) \\
&= \exp \left((1 + \phi^\tau) \mu_h + \frac{1}{2} (1 + \phi^\tau)^2 \sigma_h^2 + (1 - \phi^\tau) \mu_h + \frac{1}{2} (1 - \phi^{2\tau}) \sigma_h^2 \right) \\
&= \exp \left(2\mu_h + (1 + \phi^\tau) \sigma_h^2 \right),
\end{aligned} \tag{1.1.14}$$

il vient alors :

$$\text{Cov}[y_t^2, y_{t+\tau}^2] = \mathbb{E}[y_t^2 y_{t+\tau}^2] - \mathbb{E}[y_t^2]^2$$

$$\begin{aligned}
&= \exp\left(2\mu_h + \sigma_h^2(1 + \phi^\tau)\right) - \exp(2\mu_h + \sigma_h^2) \\
&= \exp(2\mu_h + \sigma_h^2)\left(\exp(\sigma_h^2\phi^\tau) - 1\right).
\end{aligned}$$

□

Des expressions (1.1.9) et (1.1.10), le coefficient d'aplatissement (*kurtosis*) de y_t s'écrit donc :

$$\kappa = \frac{\mathbb{E}[y_t^4]}{\mathbb{E}[y_t^2]^2} = 3 \exp(\sigma_h^2) > 3, \quad (1.1.15)$$

et de (1.1.12) et (1.1.11), la fonction d'autocorrélation du processus des rendements au carré $\{y_t^2\}$ s'exprime comme suit :

$$\text{Corr}[y_t^2, y_{t+\tau}^2] = \frac{\exp(\sigma_h^2\phi^\tau) - 1}{3 \exp(\sigma_h^2) - 1}, \quad \tau = 1, 2, \dots \quad (1.1.16)$$

Pour plus de détails sur les propriétés économétriques, les fondements théoriques et les extensions du modèle SV_0 nous référons aux travaux de Taylor (1986, 1994), Ghysels et al. (1996) et Shephard (1996).

1.2. EXTENSIONS DU MODÈLE À VOLATILITÉ STOCHASTIQUE STANDARD

Le résultat (1.1.15) nous assure que la distribution inconditionnelle du processus des rendements $\{y_t\}$ a des queues plus épaisses que celles d'une distribution normale. Cette propriété fait partie des nombreux faits stylisés que les modèles économétriques s'efforcent de répliquer. Elle provient du fait qu'il est empiriquement plus probable d'observer des rendements de plus grande amplitude que ceux qu'impliqueraient une distribution normale. La présence de sauts dans le processus des prix sous-jacents aux rendements expliquerait en partie ce phénomène (Bollerslev (1987); Kearns et Pagan (1997); Wagner et Marsh (2005)). Une illustration de ce phénomène est donnée par le graphique en haut et à droite (*Q-Q plot*)

de la figure 1.1 qui offre un aperçu de l'évolution des rendements de l'indice S&P 500 de 1999-12-31 à 2016-12-30.

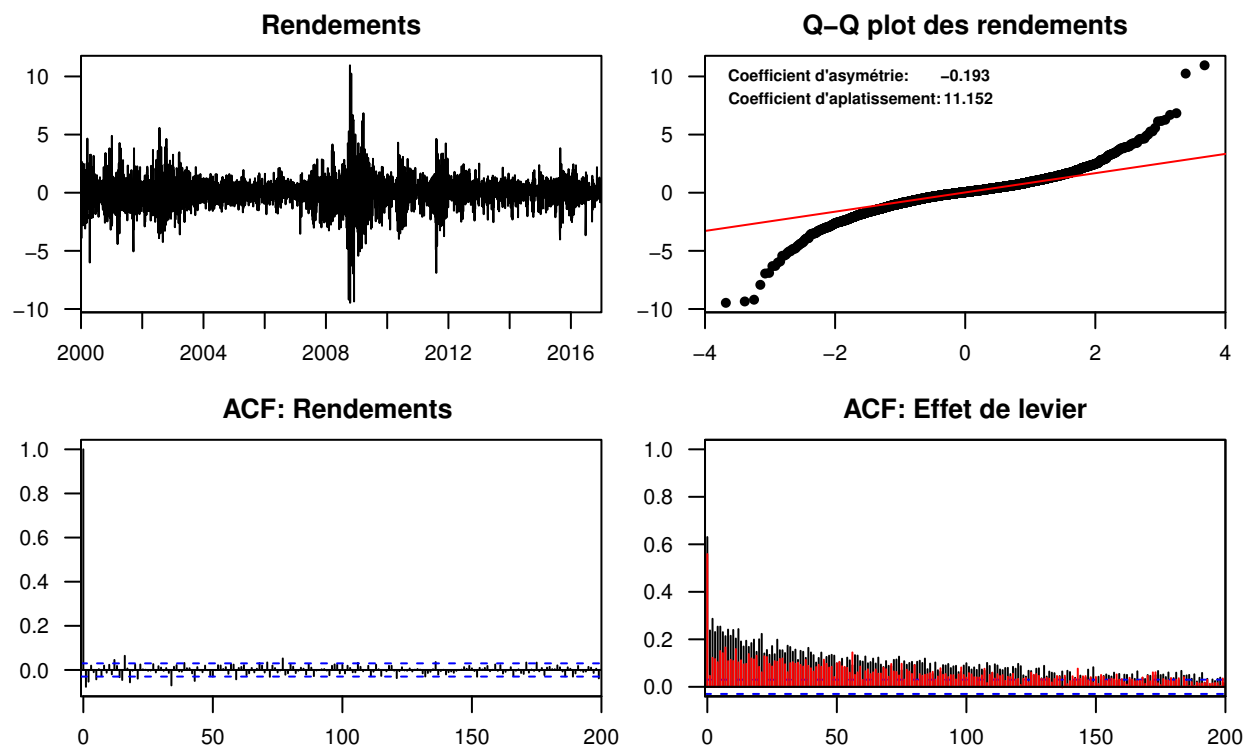


FIGURE 1.1. Rendements sur le S&P 500 au cours de la période allant du 2000-01-03 au 2016-12-30. *Haut-gauche* : évolution du rendement ; *Haut-droit* : normale Q-Q plot des rendements ; *Bas-gauche* : $\text{Corr}[y_{t+h}, y_t]$; *Bas-droit* : $\text{Corr}[|y_{t+h}|, \max(-y_t, 0)]$ en noir ; *Bas-droit* : $\text{Corr}[|y_{t+h}|, \max(y_t, 0)]$ en rouge.

Le résultat (1.1.16) suggère une décroissance exponentielle en τ de la fonction d'auto-corrélation du processus des rendements au carré $\{y_t^2\}$ qui est déterminée par le paramètre ϕ . En effet, Taylor (1986) a montré que lorsque σ_h^2 est petit et/ou ϕ est proche de 1 alors $\frac{\exp(\sigma_h^2 \phi^\tau) - 1}{3 \exp(\sigma_h^2) - 1} \approx \frac{\exp(\sigma_h^2) - 1}{3 \exp(\sigma_h^2) - 1} \phi^\tau$ (voir : Taylor (1986), pages 74-75). Ainsi, plus ce paramètre est proche de 1, ce qui est généralement le cas pour les séries financières, plus cette corrélation sera forte et permettra de mieux reproduire le phénomène de regroupement de volatilité (*volatility clustering*) qu'on observe souvent en économétrie financière.

Par contre, l'hypothèse simpliste d'indépendance entre les innovations ne permet pas au modèle SV_0 de reproduire une toute aussi importante caractéristique qu'exhibe la plupart des

séries financières à l’instar du graphique en bas et à gauche de la figure 1.1 : l’effet de levier (*leverage effect*). L’effet de levier désigne le fait qu’il existe une corrélation négative entre le rendement présent d’un actif et sa volatilité future. Économiquement, ce phénomène provient du fait qu’une baisse de la valeur d’un actif entraîne une hausse du ratio d’endettement, reflétant ainsi une hausse du risque de faillite et se traduisant par un accroissement de la volatilité future dudit actif (Black (1976); Christie (1982)).

Le modèle SV_0 ne permettant pas de répliquer tous les faits stylisés des séries financières, plusieurs extensions ont été proposées dans la littérature afin de les incorporer. Un aperçu de certaines extensions est présenté au chapitre 9 du livre de Durbin et Koopman (2012) à la page 217. Nous présentons ici deux extensions du modèle standard de volatilité stochastique à savoir : le modèle *stochastic volatility with leverage* (SVL) qui permet d’y incorporer l’effet de levier et le modèle *stochastic volatility with leverage and jumps* (SVLJ) qui, en plus de l’effet de levier, permet aussi de tenir compte des sauts dans la série de rendements observés.

1.2.1. Modèle SVL

Pour incorporer l’effet de levier dans le modèle SV_0 défini en (1.1.1a) et (1.1.1b), il suffit tout simplement d’autoriser une corrélation ρ entre les innovations ϵ_t et η_t . Cette approche a été adoptée par Harvey et Shephard (1996) et le modèle SVL qui en résulte se spécifie exactement comme la formulation (1.1.1), mais en levant l’hypothèse d’indépendance entre les innovations ϵ_t et η_t . Le vecteur des paramètres du modèle devient donc $\boldsymbol{\theta} = (\alpha, \phi, \sigma_\eta, \rho)'$.

L’estimation du paramètre supplémentaire $|\rho| \leq 1$ nous permettra donc d’évaluer empiriquement la présence et l’importance de cet effet. D’autres formulations ont été proposées dans la littérature pour incorporer l’effet de levier dans le modèle SV_0 . So et al. (2002) l’ont réalisé à l’aide d’un modèle avec seuil (*threshold*) et des innovations non corrélées. Au lieu d’une dépendance temporelle $\text{Corr}(\epsilon_t, \eta_t) = \rho$ entre les deux innovations du modèle, Jacquier et al. (2004) ont plutôt considéré $\text{Corr}(\epsilon_t, \eta_{t-1}) = \rho$. Yu (2005) a souligné dans son

étude comparative, l'incohérence de la formulation de [Jacquier et al. \(2004\)](#) avec l'hypothèse d'efficience des marchés financiers (voir : [Yu \(2005\)](#), page 168) et montré empiriquement la supériorité de la fomulation de [Harvey et al. \(1994\)](#) sur cette dernière.

Soulignons aussi que cette dépendance temporelle n'affecte pas la distribution inconditionnelle du processus $\{y_t\}$. Voir [Yu \(2005\)](#) pour plus de détails sur l'incorporation de l'effet de levier dans le modèle SV_0 .

1.2.2. Modèle SVLJ

Même si le modèle SV_0 permet d'observer un coefficient d'aplatissement plus élevé que trois, il reste par construction (produit d'une normale et d'une log-normale) limité dans la reproduction des distributions à queues très épaisses. Une première solution serait de remplacer la distribution normale de l'innovation ϵ_t par une distribution ayant une queue plus lourde à l'instar de la distribution de Student (voir : [Liesenfeld et Jung \(2000\)](#); [Chib et al. \(2002\)](#)).

Les sauts observés dans le prix des actifs financiers étant en grande partie à la source de cette croissance du coefficient d'aplatissement, il serait encore plus naturel d'incorporer des sauts dans l'équation gouvernant le processus des rendements. Nous considérons ici cette approche en autorisant, dans le modèle SVL, le processus des rendements $\{y_t\}$ de réaliser des sauts. Cette approche a été adoptée par [Pitt et al. \(2014\)](#) et permet de formuler le modèle SVLJ pour une série $y_{1:T}$ de rendements observés aux temps t ($t = 1, \dots, T$) comme suit :

$$\begin{aligned}
 y_t &= \exp\left(\frac{1}{2}h_t\right) \epsilon_t + J_t \nu_t, \\
 h_{t+1} &= \alpha + \phi h_t + \sigma_\eta \eta_t, \\
 \text{avec } \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} &\stackrel{\text{i.i.d.}}{\sim} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{1.2.1}
 \end{aligned}$$

où $\{J_t\}$ est le processus captant la présence ou non d'un saut à la date t et est modélisé comme une variable aléatoire Bernoulli de paramètre $p : J_t \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$. En présence d'un saut, c'est-à-dire, lorsque $J_t = 1$, la variable $\nu_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_J^2)$ gouverne l'intensité du saut. On suppose aussi une indépendance entre ν_t et les deux innovations ϵ_t et η_t .

Comparativement au modèle SVL, le modèle SVLJ comporte deux paramètres additionnels $0 \leq p \leq 1$ et $\sigma_J > 0$. Notons aussi que l'incorporation de l'effet de levier et des sauts dans le modèle n'affecte pas la distribution stationnaire du processus latent $\{h_t\}$.

Proposition 1.2.1. *La distribution stationnaire du processus $\{h_t\}$ pour les modèles SVL et SVLJ est identique à celle du modèle $SV_0 : h_0 \sim N(\mu_h, \sigma_h^2)$ avec μ_h et σ_h^2 donnés par les relations (1.1.3) et (1.1.4).*

DÉMONSTRATION. Ce résultat est direct en utilisant l'expression (1.1.2) pour le processus $\{h_t\}$ et l'hypothèse d'indépendance entre les processus $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ et $\xi_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. \square

La présentation des différentes approches proposées dans la littérature pour estimer les paramètres des modèles à volatilité stochastique fait l'objet de la section 1.3.

1.3. ESTIMATION DES MODÈLES À VOLATILITÉ STOCHASTIQUE

Une approche populaire pour estimer le vecteur $\boldsymbol{\theta}$ des paramètres d'un modèle économétrique sur une série de données observées $y_{1:T} := (y_1, y_2, \dots, y_T)'$ est la méthode du maximum de vraisemblance. Elle consiste à déterminer le vecteur $\hat{\boldsymbol{\theta}}$ des paramètres du modèle qui maximise la vraisemblance d'observer l'échantillon $y_{1:T}$. La fonction de vraisemblance de cet échantillon n'est rien d'autre que la densité jointe du vecteur aléatoire $Y_{1:T}$ évaluée en $y_{1:T}$ et $\boldsymbol{\theta}$ notée $p(y_{1:T} | \boldsymbol{\theta}) := L(\boldsymbol{\theta} | y_{1:T})$ et qui s'exprime comme suit :

$$p(y_{1:T} | \boldsymbol{\theta}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y_1, y_2, \dots, y_T, h_1, h_2, \dots, h_T | \boldsymbol{\theta}) dh_1 \cdots dh_T. \quad (1.3.1)$$

Dans le cadre du modèle SV_0 , cette vraisemblance s'écrit :

$$p(y_{1:T} | \boldsymbol{\theta}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y_1 | h_1, \boldsymbol{\theta}) p(h_1 | \boldsymbol{\theta}) \prod_{t=2}^T p(y_t | h_t, \boldsymbol{\theta}) p(h_t | h_{t-1}, \boldsymbol{\theta}) dh_1 \cdots dh_T, \quad (1.3.2)$$

avec $p(y_t | h_t, \boldsymbol{\theta})$ la densité de probabilité conditionnelle de y_t sachant h_t et $\boldsymbol{\theta}$ et $p(h_t | h_{t-1}, \boldsymbol{\theta})$ celle de h_t sachant h_{t-1} et $\boldsymbol{\theta}$. La détermination du maximum de vraisemblance pour ce type de modèle revient donc à optimiser une intégrale T -dimensionnelle qu'il est impossible d'évaluer directement. Une des solutions proposées dans la littérature est de contourner ce problème en estimant le modèle SV_0 par des procédures basées sur la méthode des moments généralisés (GMM). L'autre solution est de faire, soit une estimation par simulation à l'instar de [Sandmann et Koopman \(1998\)](#) et [Jacquier et al. \(1994\)](#) ou une approximation numérique de cette intégrale par une méthode déterministe comme dans [Fridman et Harris \(1998\)](#), [Bartolucci et De Luca \(2001\)](#) ou [Langrock et al. \(2012\)](#). Nous référons à l'article [Broto et Ruiz \(2004\)](#) pour la revue de la littérature sur les approches d'estimation des modèles à volatilité stochastique.

Dans cette section, nous allons faire un bref résumé des principales approches d'estimation de ce type de modèle proposées dans la littérature.

1.3.1. Méthode des moments généralisées (GMM)

[Taylor \(1986\)](#) a utilisé la méthode des moments (MM) pour estimer les paramètres du modèle SV_0 . [Melino et Turnbull \(1990\)](#) ont par la suite proposé la méthode des moments généralisées (GMM) qui consiste à trouver le vecteur des paramètres $\hat{\boldsymbol{\theta}}$ pour lequel les moments échantillonaux sont les plus proches possible des moments théoriques. Introduite dans la littérature économétrique par [Hansen \(1982\)](#), cette méthode trouve son sens dans le fait que les moments théoriques, s'exprimant à l'aide des paramètres du modèle, peuvent être estimés par leurs équivalents empiriques. [Hansen \(1982\)](#) a montré que, sous des conditions

de régularité générales, les estimateurs par la méthode GMM sont convergents et asymptotiquement normaux. Dans le cadre des modèles à volatilité stochastique, ces conditions de régularité sont assurées par la stationnarité du processus de la log-volatilité, c'est-à-dire lorsque $|\phi| < 1$. Toutefois, malgré l'avantage de cette approche d'arriver à estimer les paramètres du modèle à volatilité stochastique tout en évitant le problème d'évaluation de l'intégrale T -dimensionnelle (1.3.1), son usage sur les séries financières est peu recommandé à cause de la forte persistance ($\phi \approx 1$) qui caractérise la volatilité de ces données. En effet, [Jacquier et al. \(1994\)](#) ont trouvé dans leur étude par simulation que cette méthode produit des estimations biaisées en ce qui concerne le paramètre σ_η , en particulier lorsque ϕ est proche de 1 et que le coefficient de variation modifié de la log-volatilité, défini comme :

$$CV = \frac{\text{Var}[h_t]}{\mathbb{E}[h_t]^2} = \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right) - 1,$$

est faible. De plus, l'estimation des modèles à volatilité stochastique par la méthode GMM ne permet pas de filtrer directement les réalisations h_t du processus latent de la log-volatilité et nécessite donc l'utilisation d'une procédure de filtrage (voir : [Andersen \(1994\)](#); [Ghysels et Jasiak \(1994\)](#)).

1.3.2. Quasi-maximum de vraisemblance (QML)

L'approche du quasi-maximum de vraisemblance a été proposée par [Nelson \(1988\)](#) et [Harvey et al. \(1994\)](#) et est basée sur la réécriture du modèle SV_0 (1.1.1) dans sa forme espace-état linéaire. En effet, en prenant le logarithme de y_t^2 dans la relation (1.1.1a), on obtient l'expression suivante :

$$\log y_t^2 = h_t + \log \epsilon_t^2. \tag{1.3.3}$$

Alors, pour $\epsilon_t \sim N(0, 1)$, $\log \epsilon_t^2$ est de distribution log-chi-carré avec une moyenne donnée par $\psi(1) - \log 2 \approx 1.2704$ et de variance $\pi^2/2$ où $\psi(\cdot)$ est la fonction digamma (voir : [Abramowitz](#)

et Stegun (1972)). En posant $\xi_t = \log \epsilon_t^2 + 1.2704$, et $\omega_t = \log y_t^2$, le modèle SV_0 se réécrit comme suit :

$$\omega_t = -1.2704 + h_t + \xi_t, \quad (1.3.4a)$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \eta_t. \quad (1.3.4b)$$

où les ξ_t sont i.i.d. de moyenne nulle et de variance $\pi^2/2$.

L'idée de l'approche d'estimation par quasi-maximum de vraisemblance est de traiter les innovations ξ_t dans l'équation d'observation (1.3.4a) comme si elles étaient gaussiennes, c'est-à-dire $\xi_t \stackrel{\text{i.i.d.}}{\sim} N(0, \pi^2/2)$ et d'estimer le modèle à l'aide d'un filtre de Kalman. Pour une dérivation complète de cette approche, nous référons à l'annexe A de Sandmann et Koopman (1998), pages 297-299. La qualité de l'approximation de la distribution log-chi-carré par une distribution gaussienne est fortement liée aux valeurs des paramètres. Pour des valeurs élevées du paramètre σ_η par exemple, cette approximation sera précise puisque le terme de distribution normale h_t domine le terme non-gaussien ξ_t dans la relation (1.3.4a). À l'opposé, pour de petites valeurs de σ_η , c'est plutôt le terme non-gaussien qui est prédominant et l'approximation est alors moins précise. Ruiz (1994) a montré que les estimateurs du modèle SV_0 par QML sont convergents, asymptotiquement normaux et meilleurs que ceux de l'approche GMM. Toutefois, n'étant pas basée sur la vraie distribution de ξ_t , elle génère en échantillon fini des estimations biaisées, en particulier pour σ_η lorsque la vraie valeur de ce paramètre est petite (Jacquier et al. (1994)).

Une solution alternative basée sur le filtre de Kalman appliqué à l'approximation du logarithme du carré de l'innovation normale, a été suggérée par Kim et al. (1998). Ces derniers ont considéré une forme légèrement modifiée de la formulation espace-état du modèle SV_0 comme suit :

$$\omega_t^* = \log(y_t^2 + 0.001) = h_t + \xi_t^*, \quad (1.3.5)$$

où ξ_t^* est un mélange de K distributions normales :

$$\xi_t^* \sim \sum_{i=1}^K q_i \phi(m_i - 1.2704, \sigma_i^2), \quad (1.3.6)$$

avec q_i les poids et $\phi(\mu, \sigma^2)$ la densité de probabilité d'une normale de moyenne μ et de variance σ^2 . Les auteurs ont trouvé qu'il était possible d'obtenir une très bonne approximation de la distribution log-chi-carré par un mélange de $K = 7$ distributions normales. Les paramètres q_i , m_i et σ_i du mélange sont présentés dans la table 4 de [Kim et al. \(1998\)](#) à la page 371. Une réécriture de la relation (1.3.6) à l'aide d'une variable indicatrice s_t donne :

$$\begin{aligned} \xi_t^* \mid s_t = i &\sim \phi(m_i - 1.2704, \sigma_i^2), \\ p(s_t = i) &= q_i, \end{aligned} \quad (1.3.7)$$

les relations précédentes font ressortir que conditionnellement à s_t , ξ_t^* est sous la forme d'un modèle espace-état linéaire et gaussien.

Méthode *Monte Carlo likelihood* (MCL)

Le véritable problème de l'approche QML se trouve dans l'erreur commise en approximant la distribution log-chi-carré par une distribution normale. L'approche MCL proposée par [Sandmann et Koopman \(1998\)](#) pour estimer les paramètres du modèle SV_0 vise à corriger cette erreur d'approximation qui limite la méthode QML. La démarche proposée consiste à estimer par simulation Monte Carlo l'erreur d'approximation de la log-chi-carré par une distribution normale. Pour plus de détails sur l'algorithme et son implémentation voir [Sandmann et Koopman \(1998\)](#), pages 277-278.

1.3.3. Méthodes bayésiennes

La méthode d'inférence bayésienne vise à obtenir la distribution postérieure des paramètres au vu des données observées : $p(\boldsymbol{\theta} \mid y_{1:T})$. Par une application du théorème de Bayes, cette distribution postérieure vérifie :

$$p(\boldsymbol{\theta} \mid y_{1:T}) \propto p(\boldsymbol{\theta})p(y_{1:T} \mid \boldsymbol{\theta}), \quad (1.3.8)$$

avec $p(y_{1:T} \mid \boldsymbol{\theta})$ la fonction de vraisemblance et $p(\boldsymbol{\theta})$ la densité à priori du vecteur des paramètres $\boldsymbol{\theta}$.

Dans le cadre des modèles à volatilité stochastique il n'existe pas de solution analytique pour l'expression (1.3.8). L'idée principale de l'approche MCMC proposée par [Jacquier et al. \(1994\)](#) pour l'estimation des paramètres des modèles à volatilité stochastique est de construire récursivement une chaîne de Markov, ayant comme distribution stationnaire la distribution à posteriori (1.3.8), par le biais d'une méthode d'échantillonnage appropriée. [Jacquier et al. \(1994\)](#) ont utilisé l'échantillonnage de Gibbs à un pas qui a le désavantage de générer des trajectoires de paramètres fortement corrélées ; ceci à pour conséquence de ralentir considérablement la convergence de l'algorithme.

En plus de l'exigence en ressources computationnelles, l'autre désavantage de ces méthodes se trouve dans la complexité de les implémenter (voir : [Shephard \(2005\)](#), page 13 et [Fridman et Harris \(1998\)](#), page 285).

1.3.4. Méthodes du maximum de vraisemblance

La principale difficulté dans l'estimation des paramètres des modèles à volatilité stochastique par la méthode du maximum de vraisemblance réside dans l'incapacité d'obtenir une évaluation exacte de l'intégrale T -dimensionnelle donnée par la relation (1.3.1). Les solutions proposées dans la littérature visent donc à trouver une méthode permettant d'obtenir une

bonne approximation de cette vraisemblance afin de l’optimiser par la suite en fonction du vecteur des paramètres $\boldsymbol{\theta}$. Ces méthodes peuvent être regroupées en deux grandes catégories : les approches déterministes qui sont plus rapides et nécessitent moins de ressources informatiques et les approches non-déterministes.

Fridman et Harris (1998) ont été les premiers à proposer une approche directe de maximisation de la vraisemblance du modèle à travers une procédure d’intégration numérique récursive. Cette approche déterministe se base sur la méthodologie introduite par Kitagawa (1987) pour résoudre d’une manière générale les problèmes de filtrage non-linéaires à bruits non-gaussiens. Elle consiste à découper dans l’expression (1.3.1) chaque domaine d’intégration du processus latent $\{h_t\}$ en plusieurs morceaux et à les évaluer de manière récursive. Ce découpage de l’espace d’état du processus latent offre au modèle la même structure qu’un modèle à chaîne de Markov cachée (voir : chapitre 2). Par le biais d’une étude de simulation Monte Carlo, Fridman et Harris (1998) ont montré que cette méthode d’estimation des paramètres du modèle SV_0 est plus performante que les méthodes GMM et QML et comparable à l’approche bayésienne MCMC de Jacquier et al. (1994) et au *Simulated Maximum Likelihood* (SML) de Danielsson (1994).

Les méthodes non-déterministes exploitent les techniques de simulations pour estimer l’intégrale T -dimensionnelle (1.3.1) de la vraisemblance. Une fois qu’une bonne estimation de la vraisemblance du modèle est obtenue, cette dernière est optimisée en fonction du vecteur des paramètres. Ainsi, l’approche SML a été proposée par Danielsson (1994) dans le cadre de l’estimation des paramètres du modèle SV_0 . En supposant disposer de N réalisations $\{h_{1:T}^{(1)}, h_{1:T}^{(2)}, \dots, h_{1:T}^{(N)}\}$ de la log-volatilité latente $\{h_t\}_{t=1}^T$ issues de la distribution $p(h_{1:T} | \boldsymbol{\theta})$, une approximation de la relation (1.3.1) est donnée par :

$$L_N(\boldsymbol{\theta} | y_{1:T}) = \frac{1}{N} \sum_{i=1}^N \left[\prod_{t=1}^T p(y_t | h_t^{(i)}, \boldsymbol{\theta}) \right]. \quad (1.3.9)$$

La question centrale est donc de trouver une approche optimale pour simuler les réalisations $\{h_{1:T}^{(1)}, h_{1:T}^{(2)}, \dots, h_{1:T}^{(N)}\}$ du processus latent $\{h_t\}_{t=1}^T$. Danielsson (1994) a relevé le fait qu'une simulation naïve du processus latent sans tenir compte des observations $\{y_t\}_{t=1}^T$ conduit à des variances très élevées pour les paramètres estimés pour des séries de longueur supérieure à 200 ($T > 200$). L'auteur propose donc d'utiliser la technique d'échantillonnage préférentielle *Accelerated Gaussian Importance Sampler* de Danielsson et Richard (1993). Toutefois, il est difficile d'évaluer la qualité de cette approximation pour des valeurs du paramètre ϕ autres que zéro (voir : Jacquier et al. (1994), page 373). Aussi a-t-il été souligné par Shephard (2000) (voir : Shephard (2000), page 31) que dans le cadre du modèle SV_0 , les échantillonneurs préférentiels peuvent ne pas posséder de variance et, par conséquent, ne pas obéir au théorème central limite (TCL).

Une alternative possible à cette approche est d'utiliser une technique de simulation Monte Carlo séquentielle (*sequential Monte Carlo*) ou filtre particulaire. Le seul problème est que dans la forme classique telle qu'introduite par Gordon et al. (1993), l'estimateur obtenu pour la vraisemblance (1.3.1) est une fonction discontinue des paramètres, ce qui rend donc son optimisation difficile. Pitt et al. (2014) ont récemment proposé une approche basée sur des filtres particulaires continus pour l'estimation des paramètres des modèles SV_0 , SVL et $SVLJ$. La performance de cette approche n'a toutefois pas été comparée aux autres approches présentes dans la littérature.

Nous allons consacrer le reste du mémoire à unifier et placer ces approches déterministes de Fridman et Harris (1998); Bartolucci et De Luca (2001); Langrock et al. (2012) dans le cadre de résolution du problème de filtrage pour des modèles à chaîne de Markov cachée. Nous allons aussi montrer comment ces approches se généralisent en présence d'effet de levier et de sauts. Une comparaison de ces approches à l'approche non-déterministe basée sur les filtres particulaires de Pitt et al. (2014), dans l'approximation de la log-vraisemblance et l'estimation des paramètres des modèles à volatilité stochastique est par la suite présentée.

Nous offrons aussi une comparaison des résultats des approches *HMM* et par *filtre particulaire continu* aux méthodes MCMC de [Jacquier et al. \(1994\)](#) et MCL de [Sandmann et Koopman \(1998\)](#) pour l'estimation des paramètres du modèle à volatilité stochastique standard.

Chapitre 2

APPROCHE D'ESTIMATION DES MODÈLES À VOLATILITÉ STOCHASTIQUE BASÉE SUR LE MODÈLE À CHAÎNE DE MARKOV CACHÉE

Dans ce chapitre nous allons présenter différentes approches pour approximer la vraisemblance des modèles à volatilité stochastique. Dans un premier temps le cadre général des modèles espace-état et le problème de filtrage est introduit. Les solutions exactes et approchées de résolution de ce problème seront par la suite présentées. L'approximation de la vraisemblance des modèles à volatilité stochastique par cette méthodologie conclura le chapitre.

2.1. MODÈLES ESPACE-ÉTAT : PROBLÈME DE FILTRAGE

Les modèles à volatilité stochastique introduits dans le chapitre précédent font partie de la classe des modèles espace-état. Dans cette représentation, le problème d'estimation est résolu par un algorithme récursif qui infère ou filtre à chaque date t la densité de probabilité de la volatilité cachée du modèle sur la base des prix ou rendements observés jusqu'à t . La présente section sera consacrée à la présentation du problème de filtrage et de sa solution théorique optimale.

2.1.1. Problème de filtrage

De manière générale un modèle espace-état est composé de deux processus stochastiques ayant la représentation fonctionnelle suivante pour une série d'observation $y_{1:T}$:

$$y_t = F(h_t, \epsilon_t), \quad (2.1.1a)$$

$$h_{t+1} = G(h_t, \eta_t), \quad (2.1.1b)$$

où $\{\epsilon_t\}$ et $\{\eta_t\}$ sont des suites de processus aléatoires *i.i.d.* avec des densités de probabilité connues. F et G sont des fonctions définies et mesurables. Dans cette formulation, la variable y_t est observée et la relation (2.1.1a) est appelée **équation d'observation** ou **équation de mesure**. L'expression (2.1.1b) est appelée **équation d'état** et décrit la dynamique de la variable latente h_t qui prend ses valeurs dans l'ensemble $\mathcal{C} \subseteq \mathbb{R}$. La densité de probabilité initiale du processus caché $p(h_0)$ est spécifiée et supposée indépendante des innovations du système. Nous utilisons le terme *densité* de probabilité avec la notation $p(x)$ de manière générique pour désigner aussi bien une masse de probabilité lorsque la variable aléatoire X est discrète, qu'une densité de probabilité lorsque X est continue.

La connaissance des fonctions F et G et des densités de probabilité des innovations permettent d'avoir une reformulation probabiliste équivalente du système (2.1.1) à travers la spécification de la densité de transition $p(h_t | h_{t-1})$ et de la densité de y_t conditionnelle à h_t , $p(y_t | h_t)$:

$$h_0 \sim p(h_0), \quad (2.1.2a)$$

$$h_t | h_{0:t-1}, y_{1:t-1} \sim p(h_t | h_{t-1}), \quad (2.1.2b)$$

$$y_t | h_{1:t}, y_{1:t-1} \sim p(y_t | h_t), \quad t = 1, \dots, T. \quad (2.1.2c)$$

Le problème de filtrage (*filtering problem*) consiste à déterminer la **densité filtrée** (*filtered density*) :

$$p(h_t | y_{1:t}), \quad t = 1, \dots, T. \quad (2.1.3)$$

Les modèles à volatilité stochastique introduits au chapitre précédent constituent un exemple de modèle espace-état avec la log-volatilité h_t comme variable latente et le rendement y_t comme variable observée. Dans ce contexte, la variable latente prend ses valeurs dans \mathbb{R} . Les innovations ϵ_t et η_t sont de distributions normales indépendantes dans le cas du modèle SV_0 et corrélées dans le cas des modèles SVL et SVLJ. Pour le modèle SV_0 , les fonctions F et G sont respectivement données par $\exp(h_t/2)\epsilon_t$ et $\alpha + \phi h_t + \sigma_\eta \eta_t$.

2.1.2. Filtre optimal

Une utilisation du théorème de Bayes nous permet d'obtenir une expression pour la densité (2.1.3) en alternant successivement les étapes de prédiction (*prediction*) et de mise à jour (*update*).

Ainsi, pour $t = 1, \dots, T$, et à partir de la densité filtrée $p(h_{t-1} | y_{1:t-1})$ au temps $t - 1$, la densité prédite $p(h_t | y_{1:t-1})$ est donnée par :

Étape 1. (Prédiction) Calculer $p(h_t | y_{1:t-1})$:

$$\begin{aligned} p(h_t | y_{1:t-1}) &= \int_{\mathcal{C}} p(h_t, h_{t-1} | y_{1:t-1}) dh_{t-1}, \\ &= \int_{\mathcal{C}} p(h_t | h_{t-1}) p(h_{t-1} | y_{1:t-1}) dh_{t-1}. \end{aligned} \quad (2.1.4)$$

Suite à l'observation en t de y_t , la mise à jour permet d'obtenir la densité filtrée $p(h_t | y_{1:t})$:

Étape 2. (Mise à jour) Calculer $p(h_t | y_{1:t})$:

$$p(h_t | y_{1:t}) = \frac{p(y_t, h_t | y_{1:t-1})}{p(y_t | y_{1:t-1})},$$

$$= \frac{p(y_t | h_t)p(h_t | y_{1:t-1})}{p(y_t | y_{1:t-1})}, \quad (2.1.5)$$

avec la vraisemblance conditionnelle de y_t , $p(y_t | y_{1:t-1})$, donnée par :

$$\begin{aligned} p(y_t | y_{1:t-1}) &= \int_{\mathcal{C}} p(y_t, h_t | y_{1:t-1}) dh_t, \\ &= \int_{\mathcal{C}} p(y_t | h_t)p(h_t | y_{1:t-1}) dh_t. \end{aligned} \quad (2.1.6)$$

La log-vraisemblance du modèle peut donc être obtenue à partir de l'expression (2.1.6) :

$$\log p(y_{1:T}) = \sum_{t=1}^T \log p(y_t | y_{1:t-1}). \quad (2.1.7)$$

La récursion des étapes de prédiction (2.1.4) et de mise à jour (2.1.5) offre une solution théorique optimale au problème de filtrage (2.1.3). Toutefois, il n'est pas toujours possible d'évaluer exactement les intégrales (2.1.4) et (2.1.6), en particulier lorsque les fonctions F et G ne sont pas linéaires et/ou les innovations du système ne sont pas gaussiennes.

2.2. ÉVALUATIONS EXACTES DU FILTRE OPTIMAL

2.2.1. Filtre de Kalman

Lorsque les innovations ϵ_t et η_t sont toutes les deux gaussiennes et les fonctions F et G sont linéaires en leurs arguments, les relations (2.1.1a) et (2.1.1b) donnent lieu à un système linéaire gaussien. Dans le cas échéant, (2.1.4) et (2.1.5) sont normalement distribuées et le problème de filtrage peut être résolu de manière exacte par **le filtre de Kalman standard** (Kalman et Bucy (1961)). Il existe de nombreuses références sur les filtres de Kalman, pour une introduction nous référons au chapitre 6 de Shumway et Stoffer (2011) ou à la section *Filtre de Kalman* des notes de cours *Méthode d'inférence pour les modèles à chaîne de Markov cachée (STT6705V)* Augustyniak (2017).

Pour illustrer la récursion du filtre de Kalman, nous allons considérer ici le modèle espace-état linéaire $AR(1)$ gaussien bruité :

$$y_t = h_t + \sigma_\epsilon \epsilon_t, \quad (2.2.1a)$$

$$h_{t+1} = \phi h_t + \sigma_\eta \eta_t, \quad |\phi| < 1, \quad t = 1, \dots, T, \quad (2.2.1b)$$

$$h_0 \sim N(\hat{h}_{0|0}, P_{0|0}), \quad \text{avec } \hat{h}_{0|0} = 0 \text{ et } P_{0|0} = \frac{\sigma_\eta^2}{1 - \phi^2}, \quad (2.2.1c)$$

où $\{\epsilon_t\}$ et $\{\eta_t\}$ sont des suites indépendantes de variables aléatoires *i.i.d.* gaussiennes de moyenne zéro et de variance un. On suppose aussi que h_0 est indépendante de $\{\epsilon_t\}$ et $\{\eta_t\}$.

Proposition 2.2.1. (*Filtre de Kalman $AR(1)$ bruité*). La détermination de la densité filtrée (2.1.3) pour le système linéaire et gaussien $AR(1)$ bruité (2.2.1) s'effectue en itérant les étapes (2.1.4) et (2.1.5) pour $t = 1, \dots, T$, comme suit :

Étape 1. (*Prédiction*) Calculer $p(h_t | y_{1:t-1})$:

$$p(h_t | y_{1:t-1}) \sim N(\hat{h}_{t|t-1}, P_{t|t-1}),$$

$$\hat{h}_{t|t-1} = \phi \hat{h}_{t-1|t-1},$$

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

Étape 2. (*Mise à jour*) Calculer $p(h_t | y_{1:t})$:

$$p(h_t | y_{1:t}) \sim N(\hat{h}_{t|t}, P_{t|t}),$$

$$\hat{h}_{t|t} = \hat{h}_{t|t-1} + P_{t|t-1} \Sigma_t^{-1} (y_t - \hat{y}_t),$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} \Sigma_t^{-1} P_{t|t-1},$$

avec la vraisemblance conditionnelle de y_t , $p(y_t | y_{1:t-1})$, donnée par :

$$\begin{aligned} p(y_t | y_{1:t-1}) &\sim N(\hat{y}_t, \Sigma_t), \\ \hat{y}_t &= \hat{h}_{t|t-1}, \\ \Sigma_t &= P_{t|t-1} + \sigma_\epsilon^2. \end{aligned}$$

DÉMONSTRATION. Voir annexe (A) pour la preuve de la proposition (2.2.1). \square

2.2.2. Filtre d'Hamilton ou en espace d'états discret (*grid-based filter*)

Il est aussi possible d'obtenir une solution exacte au problème de filtrage lorsque l'espace d'états du processus caché est discret et possède un nombre fini d'état. La terminologie *chaîne de Markov cachée (HMM)* est souvent utilisée dans la littérature pour désigner un tel modèle. L'annexe (B) offre une brève introduction aux modèles à chaîne de Markov cachée.

Désignons par N , la cardinalité de l'espace d'états du processus latent $\{h_t\}$ et par $\{x^{(i)}\}_{i=1}^N$ ses valeurs possibles : $\mathcal{C} = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$. La dynamique du système est donc plus spécifiquement décrite par :

— les N éléments de la densité de probabilité initiale $p(h_0)$ (2.1.2a) :

$$p(h_0 = x^{(i)}), \quad i = 1, \dots, N \tag{2.2.2a}$$

— la densité de transition $p(h_t | h_{t-1})$ (2.1.2c) correspondant ici aux $N \times N$ éléments de la matrice de transition du processus latent :

$$p(h_t^{(i)} | h_{t-1}^{(j)}) := p(h_t = x^{(i)} | h_{t-1} = x^{(j)}), \quad i, j = 1, \dots, N, \tag{2.2.2b}$$

— et les N éléments de la densité de probabilité conditionnelle $p(y_t | h_t)$ (2.1.2b) :

$$p(y_t | h_t^{(i)}) := p(y_t | h_t = x^{(i)}), \quad i = 1, \dots, N. \tag{2.2.2c}$$

En supposant qu'au temps $t - 1$, on dispose de la densité filtrée $p(h_{t-1} | y_{1:t-1})$, composée des N probabilités d'observer chacun des états du processus latent conditionnellement aux observations disponibles à cette date et en posant :

$$p(h_{t-1}^{(i)} | y_{1:t-1}) := p(h_{t-1} = x^{(i)} | y_{1:t-1}), \quad i = 1, \dots, N,$$

nous pouvons réécrire les expressions (2.1.4) et (2.1.5) de l'algorithme itératif comme suit :

Étape 1. (Prédiction) Calculer $p(h_t^{(i)} | y_{1:t-1})$, $i = 1, \dots, N$:

$$\begin{aligned} p(h_t^{(i)} | y_{1:t-1}) &= \int_{\mathcal{C}} p(h_t = x^{(i)} | h_{t-1}) p(h_{t-1} | y_{1:t-1}) dh_{t-1}, \\ &= \sum_{j=1}^N p(h_t^{(i)} | h_{t-1}^{(j)}) p(h_{t-1}^{(j)} | y_{1:t-1}), \quad i = 1, \dots, N. \end{aligned} \quad (2.2.3)$$

Suite à l'observation en t de y_t , la mise à jour permet d'obtenir la densité filtrée $p(h_t | y_{1:t})$:

Étape 2. (Mise à jour) Calculer $p(h_t^{(i)} | y_{1:t})$, $i = 1, \dots, N$:

$$p(h_t^{(i)} | y_{1:t}) = \frac{p(y_t | h_t^{(i)}) p(h_t^{(i)} | y_{1:t-1})}{p(y_t | y_{1:t-1})}, \quad i = 1, \dots, N, \quad (2.2.4)$$

avec la vraisemblance conditionnelle de y_t , $p(y_t | y_{1:t-1})$, donnée par :

$$\begin{aligned} p(y_t | y_{1:t-1}) &= \int_{\mathcal{C}} p(y_t | h_t) p(h_t | y_{1:t-1}) dh_t, \\ &= \sum_{i=1}^N p(y_t | h_t^{(i)}) p(h_t^{(i)} | y_{1:t-1}). \end{aligned} \quad (2.2.5)$$

Une formulation vectorielle des étapes de cet algorithme itératif, appelé filtre d'Hamilton, est donnée par la proposition (2.2.2).

Proposition 2.2.2. (*Filtre d'Hamilton*). *Soit le modèle à chaîne de Markov cachée défini en (2.2.2) et posons :*

$$\vec{p}(y_t) := \left(p(y_t | h_t^{(1)}), \dots, p(y_t | h_t^{(N)}) \right)$$

$$\vec{\mathbf{w}}_{0|0} := (p(h_0^{(1)}), \dots, p(h_0^{(N)}))$$

$$\mathbf{\Gamma} := \begin{pmatrix} p(h_t^{(1)} | h_{t-1}^{(1)}) & \dots & p(h_t^{(1)} | h_{t-1}^{(N)}) \\ \vdots & p(h_t^{(i)} | h_{t-1}^{(j)}) & \vdots \\ p(h_t^{(N)} | h_{t-1}^{(1)}) & \dots & p(h_t^{(N)} | h_{t-1}^{(N)}) \end{pmatrix}.$$

Les étapes (2.2.3) et (2.2.4) de la résolution du problème de filtrage (2.1.3) se réécrivent comme suit, pour $t = 1, \dots, T$:

Étape 1. Prédiction :

$$\vec{\mathbf{w}}_{t|t-1} = \vec{\mathbf{w}}_{t-1|t-1} \mathbf{\Gamma},$$

Étape 2. Mise à jour :

$$\vec{\alpha}_t = \vec{\mathbf{w}}_{t|t-1} \odot \vec{\mathbf{p}}(y_t),$$

$$\alpha_t = \vec{\alpha}_t \mathbf{1}_N,$$

$$\vec{\mathbf{w}}_{t|t} = \frac{\vec{\alpha}_t}{\alpha_t},$$

où le symbole \odot représente le produit matriciel de Hadamard, c'est-à-dire le produit terme à terme de deux vecteurs de même dimension et $\mathbf{1}'_N = (1, \dots, 1)$.

La log-vraisemblance du modèle peut donc être obtenue par :

$$\log p(y_{1:T}) = \sum_{t=1}^T \log \alpha_t. \quad (2.2.6)$$

2.3. APPROXIMATIONS DU FILTRE OPTIMAL

Dans le contexte des systèmes non-linéaires et/ou non-gaussiens, les deux algorithmes permettant d'évaluer exactement le filtre optimal présentés dans la section 2.1.2 ne peuvent pas être exploités pour la résolution du problème. Il faudra donc proposer une solution approximant ce filtre optimal.

La première solution proposée dans la littérature consiste à linéariser le système autour de l'estimée courante et à utiliser la démarche du filtre de Kalman standard : c'est l'idée du filtre de Kalman étendu. Cette technique a l'avantage d'être simple à mettre en œuvre, mais ne produit en général pas de bons résultats.

Dans cette section, nous allons présenter les deux types d'approximations considérées dans le mémoire pour l'estimation des paramètres des modèles à volatilité stochastique : approximation numérique de la distribution filtrée par discrétisation de l'espace d'états (*filtre HMM*) et approximation numérique de la distribution filtrée par simulation Monte Carlo séquentielle (*filtre particulière*).

Sur le filtre de Kalman étendu et d'autres techniques d'approximation du filtre optimal, nous référons le lecteur au papier [Ristic et al. \(2004\)](#).

2.3.1. Filtre par approximation numérique de la distribution filtrée (*Filtre HMM*)

L'idée principale de l'approche proposée par [Kitagawa \(1987\)](#) pour la résolution du problème de filtrage dans les systèmes non-linéaires et/ou non-gaussiens est d'approcher la solution exacte du filtre optimal par approximations numériques des intégrales (2.1.4) et (2.1.6). La démarche repose sur l'évaluation des intégrandes $p(h_t | y_{1:t-1})$ et $p(h_t | y_{1:t})$ en certains points spécifiques $x^{(i)}$, $i = 1, \dots, N$ du domaine d'intégration \mathcal{C} comme suit :

Étape 1. (Prédiction) Calculer $p(h_t^{(i)} | y_{1:t-1})$, $i = 1, \dots, N$:

$$\begin{aligned} p(h_t^{(i)} | y_{1:t-1}) &= \int_{\mathcal{C}} p(h_t^{(i)} | h_{t-1}) p(h_{t-1} | y_{1:t-1}) dh_{t-1}, \\ &\approx \sum_{j=1}^N \int_{\mathcal{X}_j} p(h_t^{(i)} | h_{t-1}) p(h_{t-1} | y_{1:t-1}) dh_{t-1}, \end{aligned} \quad (2.3.1)$$

Étape 2. (Mise à jour) Calculer $p(h_t^{(i)} | y_{1:t})$, $i = 1, \dots, N$:

$$p(h_t^{(i)} | y_{1:t}) = \frac{p(y_t | h_t^{(i)}) p(h_t^{(i)} | y_{1:t-1})}{p(y_t | y_{1:t-1})}, \quad (2.3.2)$$

avec la vraisemblance conditionnelle de y_t , $p(y_t | y_{1:t-1})$, donnée par :

$$\begin{aligned} p(y_t | y_{1:t-1}) &= \int_{\mathcal{C}} p(y_t | h_t) p(h_t | y_{1:t-1}) dh_t, \\ &\approx \sum_{i=1}^N p(h_t^{(i)} | y_{1:t-1}) p(y_t | h_t^{(i)}). \end{aligned} \quad (2.3.3)$$

La log-vraisemblance du modèle peut donc être approximée par :

$$\log p(y_{1:T}) \approx \sum_{t=1}^T \log p(y_t | y_{1:t-1}). \quad (2.3.4)$$

Notons que le symbole d'approximation (\approx) dans l'expression (2.3.1) peut en réalité être remplacé par une égalité ($=$) lorsque les N intervalles $\mathcal{X}_i = (x_{i-1}, x_i)$, $i = 1, \dots, N$ forment une partition de l'espace d'états \mathcal{C} . Dans le cadre des modèles à volatilité stochastique cela suppose que les bornes inférieure (x_0) et supérieure (x_N) doivent être respectivement $-\infty$ et ∞ . Toutefois, en pratique, les bornes des intervalles extrêmes sont choisies de manière à ce que la densité de probabilité de la variable latente h_t soit négligeable en dehors des limites x_0 et x_N .

Pour évaluer numériquement l'expression (2.3.1) différentes méthodes ont été proposées par [Bartolucci et De Luca \(2001\)](#), [Clements et al. \(2004\)](#) et [Langrock et al. \(2012\)](#). Nous avons implémenté et comparé ces trois méthodes dans le présent mémoire. En désignant, dans

chaque cas, l'approximation numérique de l'intégrale $\int_{\mathcal{X}_j} p(h_t^{(i)} | h_{t-1})p(h_{t-1} | y_{1:t-1})dh_{t-1}$ par $p(h_t^{(i)} | h_{t-1}^{(j)})p(h_{t-1}^{(j)} | y_{1:t-1})$, nous avons avec $\Delta = x_i - x_{i-1}$:

-[Bartolucci et De Luca \(2001\)](#) : $x^{(i)} = x_{i-1}$ et

$$p(h_t^{(i)} | h_{t-1}^{(j)})p(h_{t-1}^{(j)} | y_{1:t-1}) \approx \Delta p(h_t = x^{(i)} | h_{t-1} = x^{(j)})p(h_{t-1} = x^{(j)} | y_{1:t-1}) \quad (2.3.5)$$

-[Langrock et al. \(2012\)](#) : $x^{(i)} = \frac{x_i + x_{i-1}}{2}$ et

$$p(h_t^{(i)} | h_{t-1}^{(j)})p(h_{t-1}^{(j)} | y_{1:t-1}) \approx p(h_t \in \mathcal{X}_i | h_{t-1} = x^{(j)})p(h_{t-1} = x^{(j)} | y_{1:t-1}) \quad (2.3.6)$$

-[Clements et al. \(2004\)](#) : $x^{(i)} = \frac{x_i + x_{i-1}}{2}$ et

$$p(h_t^{(i)} | h_{t-1}^{(j)})p(h_{t-1}^{(j)} | y_{1:t-1}) \approx \Delta p(h_t = x^{(i)} | h_{t-1} = x^{(j)})p(h_{t-1} = x^{(j)} | y_{1:t-1}). \quad (2.3.7)$$

En remplaçant cette notation dans l'expression (2.3.1), on remarque qu'approximer numériquement les intégrales (2.1.4) et (2.1.6) revient simplement à rouler le filtre d'Hamilton présenté à la section (2.2.2) sur une discrétisation de la variable latente continue h_t . En effet, en découpant l'ensemble \mathcal{C} en N intervalles $\mathcal{X}_i = (x_{i-1}, x_i)$, $i = 1, \dots, N$ et en désignant par $x^{(i)}$ le nœud ou le point représentatif de l'intervalle \mathcal{X}_i , la quantité $p(h_t^{(i)} | h_{t-1}^{(j)})$ peut être vue comme la probabilité de transition du nœud x_j au nœud x_i . Il faudrait donc veiller à ce que les probabilités de transitions $p(h_t^{(i)} | h_{t-1}^{(j)})$ soient normalisées afin de sommer à 1 pour tout j dans chacun des trois cas.

La proposition (2.3.1) offre une reformulation vectorielle des étapes de cet algorithme itératif, appelé *filtre HMM* :

Proposition 2.3.1. (*Filtre HMM*). *Soit le modèle espace-état défini en (2.1.2) avec des fonctions F et G non-linéaires et une variable d'état continue. Posons :*

$$\vec{p}(y_t) \approx \left(p(y_t | h_t = x^{(1)}), \dots, p(y_t | h_t = x^{(N)}) \right),$$

$$\vec{\mathbf{w}}_{0|0} := \left(p(h_0 = x^{(1)}), \dots, p(h_0 = x^{(N)}) \right),$$

$$\mathbf{\Gamma} := \begin{pmatrix} p(h_t = x^{(1)} | h_{t-1} = x^{(1)}) & \dots & p(h_t = x^{(1)} | h_{t-1} = x^{(N)}) \\ \vdots & p(h_t = x^{(i)} | h_{t-1} = x^{(j)}) & \vdots \\ p(h_t = x^{(N)} | h_{t-1} = x^{(1)}) & \dots & p(h_t = x^{(N)} | h_{t-1} = x^{(N)}) \end{pmatrix}.$$

Soulignons que les composantes du vecteur $\vec{\mathbf{w}}_{0|0}$ et de la matrice de transition $\mathbf{\Gamma}$ sont normalisées de sorte que les probabilités somment à 1.

Les étapes (2.3.1) et (2.3.2) de l'approximation du filtre optimal résolvant le problème de filtrage se réécrivent comme suit, pour $t = 1, \dots, T$:

Étape 1. Prédiction :

$$\vec{\mathbf{w}}_{t|t-1} := \vec{\mathbf{w}}_{t-1|t-1} \mathbf{\Gamma},$$

Étape 2. Mise à jour :

$$\begin{aligned} \vec{\alpha}_t &\approx \vec{\mathbf{w}}_{t|t-1} \odot \vec{\mathbf{p}}(y_t), \\ \alpha_t &\approx \vec{\alpha}_t \mathbf{1}_N, \\ \vec{\mathbf{w}}_{t|t} &\approx \frac{\vec{\alpha}_t}{\alpha_t}. \end{aligned}$$

La log-vraisemblance du modèle peut donc être obtenue par :

$$\log p(y_{1:T}) \approx \sum_{t=1}^T \log \alpha_t. \quad (2.3.8)$$

Notons ici que l'évaluation numérique des composantes de $\vec{\mathbf{w}}_{0|0}$ dépendra, comme pour la matrice de transition, de la méthode adoptée parmi les trois.

L'erreur d'approximation de cette vraisemblance est d'ordre $\mathcal{O}(1/N)$ (voir : [Bartolucci et De Luca \(2001\)](#), page 10). Par conséquent, il est théoriquement possible de rendre aussi

précise que l'on veut l'approximation de la vraisemblance d'un modèle espace-état non-linéaire en augmentant le nombre N de nœuds. Certaines considérations numériques sont donc nécessaires pour une bonne implémentation de cette approche de filtrage non-linéaire afin d'estimer ces types de modèles et en particulier les modèles à volatilité stochastique. Ces considérations feront en partie l'objet du chapitre 3 qui sera consacré à l'étude par simulation de la performance de cette approche.

2.3.2. Filtre particulaire

Popularisée par [Gordon et al. \(1993\)](#), l'approche des filtres particulaires vise à approcher le filtre optimal à l'aide d'un ensemble de trajectoires appelées **particules** et générées par application séquentielle des techniques d'échantillonnage préférentiel.

D'une manière générale, la densité jointe filtrée $p(h_{1:t} | y_{1:t})$ est approximée par N trajectoires pondérées évoluant selon la dynamique du processus caché. Dès qu'une nouvelle observation est disponible, elle est utilisée pour réajuster les poids des trajectoires et une technique de rééchantillonnage permet de conserver les plus cohérentes au vu de la nouvelle observation.

Cette étape de rééchantillonnage constitue un élément important de la procédure. Elle permet, d'une part, d'éviter le problème de croissance exponentielle de la variance des estimateurs obtenus par le principe d'échantillonnage préférentiel suite à l'augmentation de la dimension du vecteur simulé, et d'autre part, d'assurer un regroupement automatique des particules dans les régions d'intérêt de l'espace d'états au vu des observations.

Ces techniques sont simples à mettre en œuvre, robustes et convergent vers le filtre optimal lorsque le nombre de particules est très grand ([Del Moral \(1996\)](#)). Une importante et active littérature existe sur le sujet, on peut se référer aux livres et articles de [Cappé et al. \(2005\)](#), [Doucet et Johansen \(2009\)](#) et [Creal \(2012\)](#).

Rappel sur l'échantillonnage préférentiel

On s'intéresse à l'évaluation des intégrales de la forme :

$$I_{\varphi(X)} = \int_{\mathcal{X}} \varphi(x)p(x)dx, \quad (2.3.9)$$

où le vecteur aléatoire $X \in \mathcal{X}$ est de loi $p(x)$ et $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ est une fonction mesurable telle que $\varphi \in \mathcal{L}^1$. Supposons qu'il n'est pas facile ou possible de simuler le vecteur X à partir de la distribution $p(x)$, mais qu'on peut tirer des réalisations de X selon une distribution alternative $q(x)$, appelée proposition et dont le support inclus le support de $p(x)$:

$$p(x) > 0 \Rightarrow q(x) > 0.$$

L'intégrale (2.3.9) peut donc se réécrire :

$$\begin{aligned} I_{\varphi(X)} &= \int_{\mathcal{X}} \varphi(x)p(x)dx, \\ &= \int_{\mathcal{X}} \varphi(x) \frac{p(x)}{q(x)} q(x)dx \end{aligned} \quad (2.3.10)$$

et un estimateur de (2.3.9) est donc donné par :

$$\hat{I}_{\varphi(X)} := I_{\varphi(X)}^N = \frac{1}{N} \sum_{i=1}^N \bar{w}^{(i)} \varphi(x^{(i)}), \quad (2.3.11)$$

avec $\{x^{(i)}\}_{i=1}^N$ un échantillon *i.i.d* généré selon la distribution d'importance $q(x)$ et $\bar{w}^{(i)}$ la normalisation du poids $w^{(i)}$ de la $i^{\text{ème}}$ particule :

$$\bar{w}^{(i)} = \frac{w^{(i)}}{\sum_{i=1}^N w^{(i)}}, \quad w^{(i)} = \frac{p(x^{(i)})}{q(x^{(i)})}.$$

D'un point de vue probabiliste, l'intégrale (2.3.9) n'est rien d'autre qu'une espérance et s'écrit $I_{\varphi(X)} = \mathbb{E}[\varphi(X)]$. D'après la loi des grands nombres et le TCL, l'estimateur $I_{\varphi(X)}^N$ converge presque sûrement avec une variance asymptotique d'ordre $1/N$ vers l'intégrale $I_{\varphi(X)}$

lorsque N tend vers l'infini. Voir [Robert et Casella \(2004\)](#) pour plus de détails sur l'échantillonnage préférentiel.

Échantillonnage préférentiel séquentiel

Afin de résoudre récursivement le problème de filtrage, il faut considérer une version itérative de l'échantillonnage préférentiel (*Sequential Importance Sampling*). La distribution d'importance doit donc être choisie, non seulement pour pouvoir facilement simuler des réalisations de la variable d'intérêt, mais aussi pour permettre une utilisation itérative de la technique d'échantillonnage préférentiel. Ceci explique la nécessité de veiller à ce que le choix de la distribution d'importance $q(h_{0:t} | y_{1:t})$ soit porté sur des distributions qui peuvent s'écrire sous la forme récursive suivante :

$$q(h_{0:t} | y_{1:t}) = q(h_0) \prod_{k=1}^t q(h_k | h_{0:k-1}, y_{1:k}), \quad (2.3.12)$$

ce qui entraîne :

$$q(h_{0:t} | y_{1:t}) = q(h_t | h_{0:t-1}, y_{1:t})q(h_{0:t-1} | y_{1:t-1}). \quad (2.3.13)$$

La variable d'intérêt est dans ce contexte toute la séquence du processus latent $\{h_t\}$ jusqu'à la date t au vu des observations $y_{1:t} : \{h_{0:t} | y_{1:t}\}$.

Supposons que nous disposons à la date $t - 1$ de l'échantillon de N particules suivant :

$$\{h_{0:t-1}^{(i)}, \bar{w}_{t-1}^{(i)}\}_{i=1}^N \sim p(h_{0:t-1} | y_{1:t-1}),$$

qui doit être prolongé au temps t après l'observation y_t pour obtenir l'ensemble de N particules :

$$\{h_t^{(i)}, h_{0:t-1}^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N = \{h_{0:t}^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N \sim p(h_{0:t} | y_{1:t}).$$

Chaque nouvelle particule $\{h_{0:t}^{(i)}, \bar{w}_t^{(i)}\}$ peut donc être vue comme la particule précédente $h_{0:t-1}^{(i)}$ augmentée par la simulation $h_t^{(i)}$ de la $t^{\text{ème}}$ réalisation de la variable latente selon la distribution d'importance $q(h_t | h_{0:t-1}, y_{1:t})$ et la détermination ou la mise à jour de son poids $\bar{w}_t^{(i)}$.

En utilisant la spécification du système défini en (2.1.2) la densité jointe filtrée $p(h_{0:t} | y_{1:t})$ peut s'écrire comme suit :

$$\begin{aligned}
p(h_{0:t} | y_{1:t}) &= \frac{p(y_t, h_{0:t} | y_{1:t-1})}{p(y_t | y_{1:t-1})}, \\
&= \frac{p(y_t, h_t | y_{1:t-1}, h_{0:t-1})}{p(y_t | y_{1:t-1})} p(h_{0:t-1} | y_{1:t-1}), \\
&= \frac{p(y_t | h_t)}{p(y_t | y_{1:t-1})} p(h_t | h_{t-1}) p(h_{0:t-1} | y_{1:t-1}), \\
&\propto p(y_t | h_t) p(h_t | h_{t-1}) p(h_{0:t-1} | y_{1:t-1}). \tag{2.3.14}
\end{aligned}$$

La mise à jour des poids $\{\bar{w}_t^{(i)}\}_{i=1}^N$ est donc donnée par : $\bar{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$, où

$$\begin{aligned}
w_t^{(i)} &= \frac{p(h_{0:t}^{(i)} | y_{1:t})}{q(h_{0:t}^{(i)} | y_{1:t})}, \\
&\propto \frac{p(y_t | h_t^{(i)}) p(h_t^{(i)} | h_{t-1}^{(i)}) p(h_{0:t-1}^{(i)} | y_{1:t-1})}{q(h_t^{(i)} | h_{0:t-1}^{(i)}, y_{1:t}) q(h_{0:t-1}^{(i)} | y_{1:t-1})}, \\
&= \frac{p(h_{0:t-1}^{(i)} | y_{1:t-1})}{q(h_{0:t-1}^{(i)} | y_{1:t-1})} p(y_t | h_t^{(i)}) \frac{p(h_t^{(i)} | h_{t-1}^{(i)})}{q(h_t^{(i)} | h_{0:t-1}^{(i)}, y_{1:t})}, \\
&\propto \bar{w}_{t-1}^{(i)} p(y_t | h_t^{(i)}) \frac{p(h_t^{(i)} | h_{t-1}^{(i)})}{q(h_t^{(i)} | h_{0:t-1}^{(i)}, y_{1:t})}, \quad i = 1, \dots, N. \tag{2.3.15}
\end{aligned}$$

L'utilisation récursive des étapes (2.3.14) et (2.3.15) constitue l'algorithme d'échantillonnage préférentiel séquentiel (SIS) :

— initialiser :

$$h_0^{(i)} \text{ simuler selon la densité } p(h_0), \quad i = 1, \dots, N \text{ et fixer } \bar{w}_0^{(i)} = 1/N \tag{2.3.16a}$$

— simuler :

$$h_t^{(i)} \text{ selon la densité } q(h_t | h_{0:t-1}^{(i)}, y_{1:t}), \quad i = 1, \dots, N, \quad (2.3.16b)$$

— mettre à jour les poids des nouvelles particules avec le principe de l'échantillonnage préférentiel comme suit :

$$\bar{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}, \quad i = 1, \dots, N, \quad (2.3.16c)$$

où,

$$w_t^{(i)} := \bar{w}_{t-1}^{(i)} p(y_t | h_t^{(i)}) \frac{p(h_t^{(i)} | h_{t-1}^{(i)})}{q(h_t^{(i)} | h_{0:t-1}^{(i)}, y_{1:t})}, \quad i = 1, \dots, N.$$

— ce qui produit au final l'échantillon de N particules :

$$\{h_{0:t}^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N \sim p(h_{0:t} | y_{1:t}). \quad (2.3.16d)$$

Problème de dégénérescence des poids

La contribution de la trajectoire $h_{0:t}^{(i)}, i = 1, \dots, N$ dans l'approximation de la densité jointe filtrée $p(h_{0:t}^{(i)} | y_{1:t})$ est mesurée par son poids $\bar{w}_t^{(i)}$ normalisé. La précision de l'approximation nécessite donc qu'il n'y ait pas de particules avec des poids quasiment nuls. Ceci constitue un handicap majeur de l'algorithme (2.3.16), puisqu'après un certain nombre d'itérations tous les poids sont généralement nuls sauf un. Ce phénomène est appelé le problème de dégénérescence des poids : seulement un ou très peu de particules sont effectivement utilisées pour approximer la distribution filtrée (voir : Cappé et al. (2005), section 7.3.1).

Échantillonnage préférentiel séquentiel avec rééchantillonnage

La principale solution proposée dans la littérature pour pallier ce problème consiste à incorporer dans l'algorithme d'échantillonnage préférentiel séquentiel (2.3.16) une étape de

rééchantillonnage. Cette étape de rééchantillonnage vise à reconstituer l'ensemble des particules $\{h_{0:t}^{(i)}\}_{i=1}^N$ par l'entremise d'un tirage aléatoire avec remise sur la base des poids normalisés $\{\bar{w}_t^{(i)}\}_{i=1}^N$. Les particules ayant un faible poids normalisé auront par conséquent une grande chance d'être supprimées et remplacées par des copies des particules ayant un grand poids normalisé. Plusieurs techniques de rééchantillonnage peuvent être exploitées pour ce faire (voir : [Cappé et al. \(2005\)](#), section 7.4).

Le choix approprié de la distribution d'importance permet aussi de limiter l'effet du problème de dégénérescence des poids (voir : [Doucet et al. \(2000\)](#)).

L'observation de la relation (2.3.15) suggère que l'utilisation de $q(h_t^{(i)} | h_{0:t-1}^{(i)}, y_{1:t}) = p(h_t^{(i)} | h_{t-1}^{(i)})$ comme distribution d'importance simplifie l'expression de la mise à jour des poids des particules. Ce choix constitue en effet la solution la plus répandue et celle qui a été utilisée dans le premier filtre particulaire introduit par [Gordon et al. \(1993\)](#) pour résoudre le problème de filtrage non-linéaire et non-gaussien : *bootstrap filter*.

Proposition 2.3.2. (*Bootstrap filter*) Soit le modèle espace-état défini en (2.1.2) avec des fonctions F et G non-linéaires.

- Pour $t = 0$
 - **Initialisation** : générer $h_0^{(i)} \sim p(h_0)$, $i = 1, \dots, N$, et fixer $\bar{w}_0^{(i)} = 1/N$.
- Pour $t = 1, \dots, T$:
 - **Échantillonnage préférentiel séquentiel** :
 - (1) générer pour $i = 1, \dots, N$: $h_t^{(i)} \sim p(h_t^{(i)} | h_{t-1}^{(i)})$, et poser $h_{0:t}^{(i)} = (h_{0:t-1}^{(i)}, h_t^{(i)})$. Ceci produit les particules $\{h_{0:t}^{(i)}, \bar{w}_{t-1}^{(i)}\}_{i=1}^N \sim p(h_{0:t} | y_{1:t-1})$,
 - (2) calculer les mises à jour des poids pour $i = 1, \dots, N$, $w_t^{(i)} := \bar{w}_{t-1}^{(i)} p(y_t | h_t^{(i)})$,
 - (3) calculer les poids normalisés pour $i = 1, \dots, N$:

$$\bar{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}.$$

— **Approximation de la densité filtrée :**

à partir de l'échantillon $\{h_{0:t}^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N \sim p(h_{0:t} | y_{1:t})$, nous pouvons approximer la distribution filtrée $p(h_t | y_{1:t})$ par l'échantillon $\{h_t^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N : \{h_t^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N \sim p(h_t | y_{1:t})$.

— **Approximation de la vraisemblance conditionnelle de y_t , $p(y_t | y_{1:t-1})$:**

$$\begin{aligned}
 p(y_t | y_{1:t-1}) &= \int_{\mathcal{C}^{t+1}} p(y_t | h_{0:t}) p(h_{0:t} | y_{1:t-1}) dh_{0:t}, \\
 &= \int_{\mathcal{C}^{t+1}} p(y_t | h_t) p(h_{0:t} | y_{1:t-1}) dh_{0:t}, \\
 &\approx \sum_{i=1}^N p(y_t | h_t^{(i)}) \bar{w}_{t-1}^{(i)}, \\
 &\approx \sum_{i=1}^N w_t^{(i)}. \tag{2.3.17}
 \end{aligned}$$

— **Rééchantillonnage :**

(1) tirer avec remise N particules $\{\tilde{h}_{0:t}^{(i)}\}_{i=1}^N$ parmi les $\{h_{0:t}^{(i)}\}_{i=1}^N$ proportionnellement aux poids $\{\bar{w}_t^{(i)}\}_{i=1}^N$,

(2) poser $h_{0:t}^{(i)} = \tilde{h}_{0:t}^{(i)}$ et $\bar{w}_t^{(i)} = 1/N$ pour $i = 1, \dots, N$.

La log-vraisemblance du modèle peut donc être approximée par :

$$\log p(y_{1:T}) \approx \sum_{t=1}^T \log p(y_t | y_{1:t-1}). \tag{2.3.18}$$

L'approximation de la log-vraisemblance obtenue par le *bootstrap filter* n'est cependant pas une fonction continue des paramètres (Polson et al. (2008)), ce qui rend inefficace son utilisation pour l'estimation par maximum de vraisemblance. En effet, la discontinuité de la log-vraisemblance est une conséquence de la procédure de rééchantillonnage des particules introduite pour pallier le problème de dégénérescence des poids. Il est cependant possible de rendre cette log-vraisemblance continue en utilisant une méthode de rééchantillonnage continue. Cette version continue du filtre particulaire a été récemment proposée par Malik et Pitt

(2011) et utilisée par Pitt et al. (2014) pour l'estimation des paramètres des modèles à volatilité stochastique. On réfère le lecteur aux annexes A.1, A.2 et A.3 de Malik et Pitt (2011) pour plus de détails sur l'implémentation de cette procédure de rééchantillonnage continue. Dans le présent mémoire, nous avons utilisé et comparé le *filtre particulaire continu (CPF)* au *bootstrap filter (PF)* pour l'approximation de la log-vraisemblance de différents modèles. La version continue du filtre particulaire a été adoptée pour l'estimation des paramètres des modèles à volatilité stochastique.

2.4. APPROXIMATIONS DE LA VRAISEMBLANCE DES MODÈLES À VOLATILITÉ STOCHASTIQUE

Les expressions (1.1.1a) et (1.1.1b) montrent clairement que le système des modèles à volatilité stochastique n'est ni linéaire ni à espace d'états discret et justifie l'utilisation des solutions proposées à la section 2.3 pour la résolution du problème de filtrage pour ce type de modèle. Le but de la présente section est donc d'utiliser le *filtre HMM* et le *filtre particulaire continu* pour approximer la log-vraisemblance et estimer les paramètres des modèles à volatilité stochastique.

2.4.1. Estimation des modèles à volatilité stochastique par le *filtre HMM*

La proposition (2.3.1) offre un algorithme général pour approximer la log-vraisemblance d'un modèle espace-état non-linéaire et/ou non-gaussien. Contrairement à l'approche par filtre particulaire, l'approximation obtenue par le *filtre HMM* est par construction continue et différentiable en ses paramètres. L'utilisation de cette approche pour estimer les modèles à volatilité stochastique revient donc à spécifier le vecteur de probabilité initial $\vec{w}_{0|0}$, le vecteur de probabilité de vraisemblance conditionnelle $\vec{p}(y_t)$ et la matrice des probabilités de transition $\mathbf{\Gamma}_t$ à chaque date $t = 1, \dots, N$.

Une fois qu'une approximation numérique de la log-vraisemblance est obtenue, il ne restera qu'à faire appel à une routine d'optimisation numérique (comme *nlm*, *optim* ou *solnp* dans R) pour obtenir une estimation par maximum de vraisemblance des paramètres du modèle.

- Estimation du modèle SV_0

Les expressions (2.4.1) présentent une reformulation probabiliste du modèle à volatilité stochastique standard (1.1.1) :

$$p(h_0) \sim N(u_h, \sigma_h^2), \quad (2.4.1a)$$

$$p(h_t | h_{t-1}) \sim N(\alpha + \phi h_{t-1}, \sigma_\eta^2), \quad (2.4.1b)$$

$$p(y_t | h_t) \sim N(0, \exp(h_t)), \quad t = 1, \dots, T, \quad (2.4.1c)$$

ce qui permet donc d'écrire :

$$\begin{aligned} \vec{p}(y_t) &= \left(p(y_t | h_t^{(1)}), \dots, p(y_t | h_t^{(N)}) \right), \\ &= \left(\{y_t; 0, \exp(x^{(1)})\}, \dots, \{y_t; 0, \exp(x^{(N)})\} \right), \end{aligned} \quad (2.4.2)$$

$$\begin{aligned} \vec{w}_{0|0} &= \left(p(h_0^{(1)}), \dots, p(h_0^{(N)}) \right), \\ &= \left(\{x^{(1)}; \mu_h, \sigma_h^2\}, \dots, \{x^{(N)}; \mu_h, \sigma_h^2\} \right), \end{aligned} \quad (2.4.3)$$

$$\mathbf{\Gamma}_t = \begin{pmatrix} p(h_t^{(1)} | h_{t-1}^{(1)}) & \dots & p(h_t^{(1)} | h_{t-1}^{(N)}) \\ \vdots & p(h_t^{(i)} | h_{t-1}^{(j)}) & \vdots \\ p(h_t^{(N)} | h_{t-1}^{(1)}) & \dots & p(h_t^{(N)} | h_{t-1}^{(N)}) \end{pmatrix},$$

$$\text{où } p(h_t^{(i)} | h_{t-1}^{(j)}) = \{x^{(i)}; \alpha + \phi h_{t-1}, \sigma_\eta^2\}, \quad i, j = 1, \dots, N, \quad (2.4.4)$$

avec la notation :

$$\{x; \mu, \sigma^2\} := \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}.$$

Rappelons que les composantes du vecteur $\vec{\omega}_{0|0}$ et de la matrice de transition Γ_t dépendent de la méthode d'évaluation numérique adoptée parmi les trois (2.3.5, 2.3.6 et 2.3.7) et sont normalisées de sorte que les probabilités somment à 1.

Cette approche d'estimation des paramètres du modèle SV_0 est similaire à celles utilisées dans les articles [Fridman et Harris \(1998\)](#), [Bartolucci et De Luca \(2001\)](#) et [Langrock et al. \(2012\)](#), mais ceux-ci n'ont pas placé la démarche dans le cadre plus général des modèles espace-état. La résolution du problème de filtrage et l'approximation de la log-vraisemblance du modèle SV_0 s'obtiennent en appliquant le *filtre HMM* (2.3.1) avec les vecteurs $\vec{\omega}_{0|0}$, $\vec{p}(y_t)$ et la matrice de transition Γ_t donnés respectivement par les expressions (2.4.2), (2.4.3) et (2.4.4).

- Estimation des modèles SVL et SVLJ

Il est aisé de remarquer que la log-vraisemblance du modèle s'approximerait aussi sans difficulté additionnelle si on avait supposé une distribution à queue lourde (*student* par exemple) pour les rendements dans l'expression (1.1.1a). En effet, il suffirait d'adapter le vecteur $\vec{p}(y_t)$ en conséquence. De même, l'intégration de l'effet de levier (SVL) et des sauts (SVLJ) dans le modèle peuvent se faire en ajustant la matrice de transition Γ_t non-homogène et/ou le vecteur de probabilité de vraisemblance conditionnelle $\vec{p}(y_t)$. Au mieux de nos connaissances, l'utilisation d'une telle approche pour l'estimation du modèle SVLJ n'a pas encore été proposée dans la littérature.

Nous présentons ici les ajustements nécessaires aux expressions (2.4.2) et (2.4.4) pour estimer le modèle SVLJ (1.2.1) avec le *filtre HMM*. Notons que l'expression (2.4.3) reste inchangée pour les modèles SVL et SVLJ, ces dernières ayant la même distribution stationnaire que le modèle SV_0 , voir proposition 1.2.1. De plus, les expressions du vecteur $\vec{p}(y_t)$ et

de la matrice de transition Γ_t peuvent être obtenues pour le modèle SVL à partir de celles du modèle SVLJ en imposant la contrainte $p = 0$.

Pour rappel, le modèle SVLJ est formulé comme suit :

$$y_t = \exp\left(\frac{1}{2}h_t\right)\epsilon_t + J_t\nu_t, \quad (2.4.5a)$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \rho \epsilon_t + \sigma_\eta \sqrt{1 - \rho^2} \xi_t, \quad t = 1, \dots, T, \quad (2.4.5b)$$

où $\{\epsilon_t\}$ et $\{\xi_t\}$ sont deux suites de processus aléatoires *i.i.d.* de distributions normales centrées en zéro et de variances égales à l'unité ($\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ et $\xi_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$) et $\{J_t\}$ est le processus captant la présence ou non de saut à la date t . Il est modélisé comme une variable aléatoire Bernoulli de paramètre p : $J_t \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$. En présence de saut, c'est-à-dire, lorsque $J_t = 1$, la variable $\nu_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_j^2)$ gouverne l'intensité du saut. On suppose aussi une indépendance entre ν_t et les deux innovations ϵ_t et ξ_t .

La densité de transition du processus latent h_t dépend donc de la variable ϵ_{t-1} . Conditionnellement à (h_{t-1}, y_{t-1}) , ϵ_{t-1} est distribuée selon un mélange comme l'indique l'équation :

$$\epsilon_{t-1} = \begin{cases} y_{t-1}e^{-\frac{1}{2}h_{t-1}} & \text{si } J_{t-1} = 0, \\ (y_{t-1} - \nu_{t-1})e^{-\frac{1}{2}h_{t-1}} & \text{si } J_{t-1} = 1. \end{cases}$$

La densité de probabilité conditionnelle de l'innovation est donnée par :

$$p(\epsilon_{t-1} | h_{t-1}, y_{t-1}) = \sum_{j=0}^1 p(\epsilon_{t-1} | J_{t-1} = j, h_{t-1}, y_{t-1})p(J_{t-1} = j | h_{t-1}, y_{t-1}). \quad (2.4.6)$$

Avec :

$$p(J_{t-1} = 0 | h_{t-1}, y_{t-1}) = 1 - p(J_{t-1} = 1 | h_{t-1}, y_{t-1}), \quad (2.4.7)$$

$$\begin{aligned} p(J_{t-1} = 1 | h_{t-1}, y_{t-1}) &= \frac{p(y_{t-1} | h_{t-1}, J_{t-1} = 1)p(J_{t-1} = 1)}{\sum_{j=0}^1 p(y_{t-1} | h_{t-1}, J_{t-1} = j)p(J_{t-1} = j)}, \quad (2.4.8) \\ &= \frac{p \times \{y_{t-1}; 0, \exp(h_{t-1}) + \sigma_j^2\}}{p \times \{y_{t-1}; 0, \exp(h_{t-1}) + \sigma_j^2\} + (1 - p) \times \{y_{t-1}; 0, \exp(h_{t-1})\}}, \end{aligned}$$

et en notant $\delta_{x_0}(x)$ la distribution de Dirac au point x_0 ,

$$p(\epsilon_{t-1} \mid J_{t-1} = 0, h_{t-1}, y_{t-1}) = \delta_{y_{t-1}e^{-\frac{1}{2}h_{t-1}}}(\epsilon_{t-1}), \quad (2.4.9)$$

$$\begin{aligned} p(\epsilon_{t-1} \mid J_{t-1} = 1, h_{t-1}, y_{t-1}) &\propto p(y_{t-1} \mid J_{t-1} = 1, h_{t-1}, \epsilon_{t-1})p(\epsilon_{t-1} \mid J_{t-1} = 1, h_{t-1}), \\ &= p(y_{t-1} \mid J_{t-1} = 1, h_{t-1}, \epsilon_{t-1})p(\epsilon_{t-1}), \\ &= \{y_{t-1}; \epsilon_{t-1} \exp(h_{t-1}/2), \sigma_J^2\} \times \{\epsilon_{t-1}; 0, 1\}. \end{aligned} \quad (2.4.10)$$

Il vient alors, voir (A.2), que :

$$p(\epsilon_{t-1} \mid J_{t-1} = 1, h_{t-1}, y_{t-1}) = \{\epsilon_{t-1}; \mu_{\epsilon_{t-1}}, \sigma_{\epsilon_{t-1}}^2\}, \quad (2.4.11)$$

avec

$$\mu_{\epsilon_{t-1}} = \frac{y_{t-1} \exp(h_{t-1}/2)}{\exp(h_{t-1}) + \sigma_J^2}, \quad \sigma_{\epsilon_{t-1}}^2 = \frac{\sigma_J^2}{\exp(h_{t-1}) + \sigma_J^2}.$$

Avec les notations $q_t := \mathbb{P}(J_{t-1} = 1 \mid h_{t-1}, y_{t-1})$ et :

$$p(h_t \mid J_{t-1} = j, h_{t-1}, y_{t-1}) = \begin{cases} \left\{ h_t; \alpha + \phi h_{t-1} + \sigma_\eta \rho y_{t-1} \exp(-h_{t-1}/2), \right. \\ \left. \sigma_\eta^2(1 - \rho^2) \right\}, j = 0, \\ \left\{ h_t; \alpha + \phi h_{t-1} + \sigma_\eta \rho \frac{y_{t-1} \exp(h_{t-1}/2)}{\exp(h_{t-1}) + \sigma_J^2}, \right. \\ \left. \frac{\sigma_\eta^2 \rho^2 \sigma_J^2}{\exp(h_{t-1}) + \sigma_J^2} + \sigma_\eta^2(1 - \rho^2) \right\}, j = 1, \end{cases}$$

nous pouvons donc exprimer la densité de transition de la variable latente comme un mélange de deux densités comme suit :

$$p(h_t \mid h_{t-1}, y_{t-1}) = q_t \times p(h_t \mid J_{t-1} = 1, h_{t-1}, y_{t-1}) + (1 - q_t) \times p(h_t \mid J_{t-1} = 0, h_{t-1}, y_{t-1}). \quad (2.4.12)$$

Remarque: La densité de transition dépend de y_{t-1} et est donc non-homogène.

Aussi, de la relation (2.4.5a), la distribution conditionnelle $p(y_t | h_t)$ s'exprime comme un mélange :

$$p(y_t | h_t) = p \times \{y_t; 0, \exp(h_t) + \sigma_J^2\} + (1 - p) \times \{y_t; 0, \exp(h_t)\}. \quad (2.4.13)$$

Ce qui entraîne les expressions (2.4.14) comme représentation probabiliste du modèle SVLJ :

$$h_0 \sim N(u_h, \sigma_h^2), \quad (2.4.14a)$$

$$h_t | h_{t-1}, y_{t-1} \sim p(h_t | h_{t-1}, y_{t-1}), \quad (2.4.14b)$$

$$y_t | h_t \sim p(y_t | h_t), \quad t = 1, \dots, T, \quad (2.4.14c)$$

où $p(h_t | h_{t-1}, y_{t-1})$ et $p(y_t | h_t)$ sont respectivement données par les expressions (2.4.12) et (2.4.13), et nous permet d'écrire pour le modèle SVLJ :

$$\vec{p}(y_t) = \left(p(y_t | h_t^{(1)}), \dots, (y_t | h_t^{(N)}) \right), \text{ avec} \quad (2.4.15)$$

$$p(y_t | h_t = x^{(i)}) = p \times \{y_t; 0, \exp(x^{(i)}) + \sigma_J^2\} + (1 - p) \times \{y_t; 0, \exp(x^{(i)})\}, \quad i = 1, \dots, N.$$

$$\mathbf{\Gamma}_t = \begin{pmatrix} p(h_t^{(1)} | h_{t-1}^{(1)}, y_{t-1}) & \dots & p(h_t^{(1)} | h_{t-1}^{(N)}, y_{t-1}) \\ \vdots & p(h_t^{(i)} | h_{t-1}^{(j)}, y_{t-1}) & \vdots \\ p(h_t^{(N)} | h_{t-1}^{(1)}, y_{t-1}) & \dots & p(h_t^{(N)} | h_{t-1}^{(N)}, y_{t-1}) \end{pmatrix}. \quad (2.4.16)$$

où pour $i, j = 1, \dots, N$:

$$\begin{aligned} p(h_t = x^{(i)} | h_{t-1} = x^{(j)}, y_{t-1}) &= q_t \times p(h_t^{(j)} | J_{t-1} = 1, h_{t-1}^{(i)}, y_{t-1}) \\ &\quad + (1 - q_t) \times p(h_t^{(j)} | J_{t-1} = 0, h_{t-1}^{(i)}, y_{t-1}). \end{aligned}$$

Nous pouvons donc résoudre le problème de filtrage et approximer la log-vraisemblance du modèle SVLJ (ou du modèle SVL avec la contrainte $p = 0$) en appliquant le *filtre HMM* (2.3.1) avec les vecteurs $\vec{w}_{0|0}$, $\vec{p}(y_t)$ et la matrice de transition Γ_t donnés respectivement par les expressions (2.4.3), (2.4.15) et (2.4.16).

2.4.2. Estimation des modèles à volatilité stochastique par *filtre particulaire continu*

L'implémentation de la proposition (2.3.2) pour l'approximation de la log-vraisemblance des modèles à volatilité stochastique est directe à condition qu'il soit possible de simuler à partir de la densité de transition $p(h_t | h_{t-1}, y_{t-1})$ de la variable latente et d'évaluer la vraisemblance conditionnelle $p(y_t | h_t)$ du modèle. En effet, ces deux densités sont indispensables pour prolonger de période en période l'ensemble de particules $\{h_{0:t}^{(i)}, \bar{w}_t^{(i)}\}_{i=1}^N, t = 1, \dots, T$.

Nous allons spécifier dans cette section les densités à utiliser pour résoudre le problème de filtrage et approximer la log-vraisemblance dans le cadre du modèle SVLJ par *filtre particulaire continu*.

- (1) Les N particules initiales du filtre sont simulées avec la distribution $p(h_0)$ donnée par l'expression (2.4.14a);
- (2) les N réalisations de la variable latente $h_t^{(i)}$ sont simulées avec la densité de transition $p(h_t | h_{t-1}, y_{t-1})$ définie par la relation (2.4.12);
- (3) la densité pour mettre à jour les poids des particules $p(y_t | h_t^{(i)})$ est donnée par la relation (2.4.13);
- (4) l'étape de rééchantillonnage des particules est réalisée avec la technique de rééchantillonnage continu introduite par Malik et Pitt (2011) (voir : Malik et Pitt (2011), annexes A.1, A.2 et A.3);
- (5) finalement, la log-vraisemblance du modèle est approximée par l'expression (2.3.18).

L'approximation de la log-vraisemblance ainsi que l'estimation des paramètres des modèles SV_0 et SVL peuvent s'en déduire en imposant respectivement $(\rho = 0, p = 0)$ et $(p = 0)$ comme contraintes aux expressions du modèle SVLJ.

Nous consacrerons les prochains chapitres à l'évaluation de la qualité et la comparaison de ces deux approches pour l'approximation de la log-vraisemblance et l'estimation des paramètres des modèles à volatilité stochastique sur des données simulées et empiriques.

Chapitre 3

ÉTUDES PAR SIMULATIONS DE LA QUALITÉ DE L'APPROXIMATION DE LA VRAISEMBLANCE

Ce chapitre sera consacré à l'étude de la qualité de l'approximation de la log-vraisemblance des modèles espace-état par les approches du *filtre HMM* et du *filtre particulaire*. Nous considérons dans un premier temps l'approximation de la log-vraisemblance du modèle $AR(1)$ gaussien bruité introduit à la section 2.2.1. Par la suite, nous comparons les approximations de la log-vraisemblance des modèles à volatilité stochastique SV_0 , SVL et $SVLJ$ sur la série du *S&P 500* par le *filtre HMM* et le *filtre particulaire*.

3.1. CONSIDÉRATIONS NUMÉRIQUES D'IMPLÉMENTATION

Certaines précautions pratiques nécessitent d'être prises en considération pour une bonne implémentation de ces deux approches et en particulier de l'approche du *filtre HMM*. Le seul élément essentiel dans l'implémentation du filtre particulaire est le choix du nombre M de particules à simuler. Nous avons évalué la qualité du *filtre particulaire continu* et du *bootstrap filter* pour différents nombres de particules M : 500, 1000, 10^4 , 10^5 et 5×10^5 .

En ce qui concerne le *filtre HMM*, l'efficacité de l'algorithme pour l'approximation de la log-vraisemblance et l'estimation des paramètres des modèles à volatilité stochastique dépendent fortement du nombre N d'intervalles de discrétisation ou de nœuds et du support

d'intégration considérés pour le processus latent $\{h_t\}$. D'une manière générale, il faut fixer les bornes inférieure (x_0) et supérieure (x_N) de sorte que, d'une part, la densité de probabilité des valeurs en dehors du support d'intégration soit négligeable, et d'autre part, pas trop large afin que les intervalles soient suffisamment petits sans quoi les nœuds ne seraient plus représentatifs de ces derniers.

Une solution facile consisterait à choisir les limites x_0 et x_N les plus extrêmes possibles et augmenter le nombre N de nœuds dans l'approximation afin d'avoir des intervalles suffisamment petits. Toutefois, cette approche est inefficace d'un point de vue numérique du filtre à cause de l'augmentation du temps de calcul de la log-vraisemblance qui est d'ordre $\mathcal{O}(TN^2)$.

Nous avons exploré la sensibilité de l'approximation de la log-vraisemblance pour différentes valeurs de x_0 , x_N et N . Les valeurs suivantes : 30, 50, 70, 100, 300 et 500 ont été considérées pour N .

Les $N + 1$ bornes des N intervalles sont données par :

$$x_i = x_0 + i \times \Delta, \quad \text{avec} \quad \Delta = \frac{x_N - x_0}{N}, \quad i = 0, 1, \dots, N.$$

Les valeurs des limites inférieure (x_0) et supérieure (x_N) sont obtenues à l'aide de la distribution stationnaire du processus latent :

$$-(x_0 - \mu_h) = x_N - \mu_h := k \times \sigma_h, \quad k = 3, 4, 5.$$

La valeur de k a été fixée à trois dans [Fridman et Harris \(1998\)](#) et cinq dans [Bartolucci et De Luca \(2001\)](#). Aussi, [Fridman et Harris \(1998\)](#) ont suggéré de choisir le nombre d'intervalles de discrétisation de sorte que l'étendue $\Delta = x_i - x_{i-1} \leq \sigma_\eta/2$, $i = 1, \dots, N$.

3.2. QUALITÉ DE L'APPROXIMATION DE LA LOG-VRAISEMBLANCE DU MODÈLE $AR(1)$ BRUITÉ

Nous présentons dans cette section les résultats des expériences de simulations menées dans le but d'évaluer la qualité d'approximation de la log-vraisemblance dans le cadre du modèle $AR(1)$ bruité. L'intérêt du choix du modèle $AR(1)$ bruité réside dans la possibilité d'évaluer de manière exacte sa log-vraisemblance avec le filtre de Kalman. Nous adoptons la spécification (2.2.1) du modèle et exploitons la proposition 2.2.1 pour évaluer de façon exacte sa log-vraisemblance avec le filtre de Kalman. La qualité de chacun des deux filtres approximatifs peut donc être évaluée en utilisant le filtre de Kalman comme référence.

3.2.1. Étude Monte Carlo sur le modèle $AR(1)$ bruité

Dans cette optique, nous avons simulé 500 réalisations du modèle $AR(1)$ bruité, de longueur $T = 1000$ chacune. Les paramètres utilisés pour simuler ces séries ont été fixés à :

$$\phi = 0.98, \sigma_\epsilon = 0.4 \text{ et } \sigma_\eta = 0.2.$$

Pour chacune des séries, les approximations de la log-vraisemblance $\hat{\ell}_i, i = 1, \dots, 500$, par *filtre HMM* sont calculées. Les points ou nœuds considérés pour ces approximations dépendent de la méthode utilisée et sont donnés par les expressions (2.3.5), (2.3.6) et (2.3.7) respectivement pour les approches HMM_1 (Bartolucci et De Luca (2001)), HMM_2 (Lan-grock et al. (2012)) et HMM_3 (Clements et al. (2004)). Les mêmes valeurs des paramètres ($\phi = 0.98, \sigma_\epsilon = 0.4$ et $\sigma_\eta = 0.2$) utilisées pour simuler les séries ont été utilisées pour évaluer les log-vraisemblances.

Chacune de ces approximations est par la suite comparée à la valeur exacte produite par le filtre de Kalman $\ell_i, i = 1, \dots, 500$.

La moyenne ($\bar{\Delta}$), ainsi que l'écart-type (Δ_{sd}) des erreurs relatives d'approximation

$$\Delta_i = \frac{\ell_i - \hat{\ell}_i}{|\ell_i|}, \quad i = 1, \dots, 500$$

ont été calculés pour apprécier la qualité de ces approches.

La même analyse a été réalisée pour les méthodes par *filtre particulaire continu* et par *bootstrap filter*. Les approches par filtres particuliers étant basées sur la simulation, les approximations en sont donc tributaires et varient d'une évaluation à une autre sur la même série. Ce problème se pose surtout lorsque le nombre de particules considéré est petit. Pour atténuer l'effet de l'aléa pour cette approche, nous avons réalisé, pour chacune des 500 séries, 100 approximations sur la même série et considéré leur moyenne comme approximation $\hat{\ell}_i$ de la log-vraisemblance de la série $i, i = 1, \dots, 500$.

3.2.2. Résultats des études sur l'approximation de la log-vraisemblance du modèle $AR(1)$ bruité

Les tableaux 3.1 et 3.2 résument les résultats pour les filtres HMM et le tableau 3.3 pour les filtres particuliers. Il en ressort que ces approches peuvent approximer, avec une très grande précision, la log-vraisemblance d'un modèle espace-état linéaire et gaussien, en l'occurrence le modèle $AR(1)$ gaussien bruité.

Ce niveau de précision dépend toutefois du paramètre k et de la méthode adoptée pour l'approximation numérique de la solution du filtre optimal en ce qui concerne les filtres HMM et du nombre de particules M pour les filtres particuliers. En effet, lorsque $k = 5$, l'erreur d'approximation est quasi-nulle même pour un petit nombre d'intervalles discrets $N = 30$. Par contre, indépendamment du nombre d'intervalles N , les erreurs d'approximations sont plus élevées lorsque le domaine d'intégration considéré pour la variable latente est restreint à $k = 3$ (tableaux 3.1, 3.2).

TABLEAU 3.1. Qualité d'approximation de la log-vraisemblance d'un modèle $AR(1)$ bruité par *filtre HMM*, selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k) (Début).

<i>Filtre</i>	$\overline{\log\text{-vrais}}$	$100 \times \bar{\Delta}$	$100 \times \Delta_{sd}$	<i>Temps de calcul(s)</i>
<i>Kalman</i>	745.0415			0.011
<i>k</i>	<i>N = 30</i>			
HMM ₁	747.6242	-0.3455	1.4626	0.015
3 HMM ₂	747.7014	-0.3576	1.4577	0.016
HMM ₃	747.4943	-0.3281	1.4470	0.016
HMM ₁	745.0567	-0.0020	0.0500	0.015
4 HMM ₂	745.6530	-0.0850	0.1688	0.016
HMM ₃	745.0685	-0.0036	0.0832	0.016
HMM ₁	745.0735	0.0005	0.0044	0.017
5 HMM ₂	746.3638	0.0250	0.0294	0.018
HMM ₃	745.0660	0.0003	0.0044	0.017
<i>k</i>	<i>N = 50</i>			
HMM ₁	747.5131	-0.3306	1.4189	0.024
3 HMM ₂	747.4969	-0.3291	1.4319	0.025
HMM ₃	747.4703	-0.3249	1.4289	0.027
HMM ₁	745.0605	-0.0025	0.0610	0.025
4 HMM ₂	745.1406	-0.0144	0.1002	0.026
HMM ₃	745.069	-0.0037	0.0830	0.026
HMM ₁	745.0415	0.0000	0.0001	0.027
5 HMM ₂	745.2267	-0.0265	0.0851	0.027
HMM ₃	745.0414	0.0000	0.0002	0.028
<i>k</i>	<i>N = 70</i>			
HMM ₁	747.4831	-0.3266	1.4108	0.035
3 HMM ₂	747.4279	-0.3251	1.4251	0.034
HMM ₃	747.463	-0.3230	1.4237	0.037
HMM ₁	745.0624	-0.0028	0.0665	0.038
4 HMM ₂	745.0837	-0.0062	0.0878	0.038
HMM ₃	745.0689	-0.0037	0.0828	0.039
HMM ₁	745.0415	0.0000	0.0001	0.042
5 HMM ₂	745.085	-0.0067	0.0445	0.039
HMM ₃	745.0414	0.0000	0.0002	0.045

TABLEAU 3.2. Qualité d'approximation de la log-vraisemblance d'un modèle $AR(1)$ bruité par *filtre HMM*, selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k) (Fin).

<i>Filtre</i>	$\overline{\log-vrais}$	$100 \times \bar{\Delta}$	$100 \times \Delta_{sd}$	<i>Temps de calcul(s)</i>
<i>Kalman</i>	745.0415			0.011
<i>k</i>	<i>N = 100</i>			
HMM ₁	747.4677	-0.3246	1.4088	0.065
3 HMM ₂	747.4603	-0.3237	1.4215	0.066
HMM ₃	747.4590	-0.3234	1.4209	0.065
HMM ₁	745.0641	-0.0030	0.0710	0.068
4 HMM ₂	745.0702	-0.0041	0.0840	0.066
HMM ₃	745.0689	-0.0036	0.0827	0.067
HMM ₁	745.0415	0.0000	0.0001	0.067
5 HMM ₂	745.0487	-0.0014	0.0221	0.068
HMM ₃	745.0414	0.0000	0.0002	0.067
<i>k</i>	<i>N = 300</i>			
HMM ₁	747.4573	-0.3232	1.4111	0.218
3 HMM ₂	747.4561	-0.3230	1.4190	0.210
HMM ₃	747.4561	-0.3231	1.4194	0.212
HMM ₁	745.0663	-0.0033	0.0765	0.229
4 HMM ₂	745.0681	-0.0036	0.0827	0.221
HMM ₃	745.0688	-0.0036	0.0826	0.243
HMM ₁	745.0414	0.0000	0.0002	0.270
5 HMM ₂	745.0406	0.0000	0.0056	0.235
HMM ₃	745.0414	0.0000	0.0002	0.278
<i>k</i>	<i>N = 500</i>			
HMM ₁	747.455	-0.3229	1.4148	1.238
3 HMM ₂	747.4553	-0.3240	1.4183	1.241
HMM ₃	747.4553	-0.3229	1.4183	1.295
HMM ₁	745.0678	-0.0035	0.0801	1.385
4 HMM ₂	745.0686	-0.0036	0.0826	1.376
HMM ₃	745.0688	-0.0036	0.0825	1.389
HMM ₁	745.0414	0.0000	0.0002	1.584
5 HMM ₂	745.0412	0.0000	0.0009	1.501
HMM ₃	745.0414	0.0000	0.0002	1.602

Note : HMM_1 , HMM_2 et HMM_3 correspondent respectivement aux approximations spécifiées par les expressions (2.3.5), (2.3.6) et (2.3.7). Le temps de calcul représente le temps moyen (en seconde) mis par l'approche pour évaluer une fois la log-vraisemblance pour une série de longueur $T = 1000$.

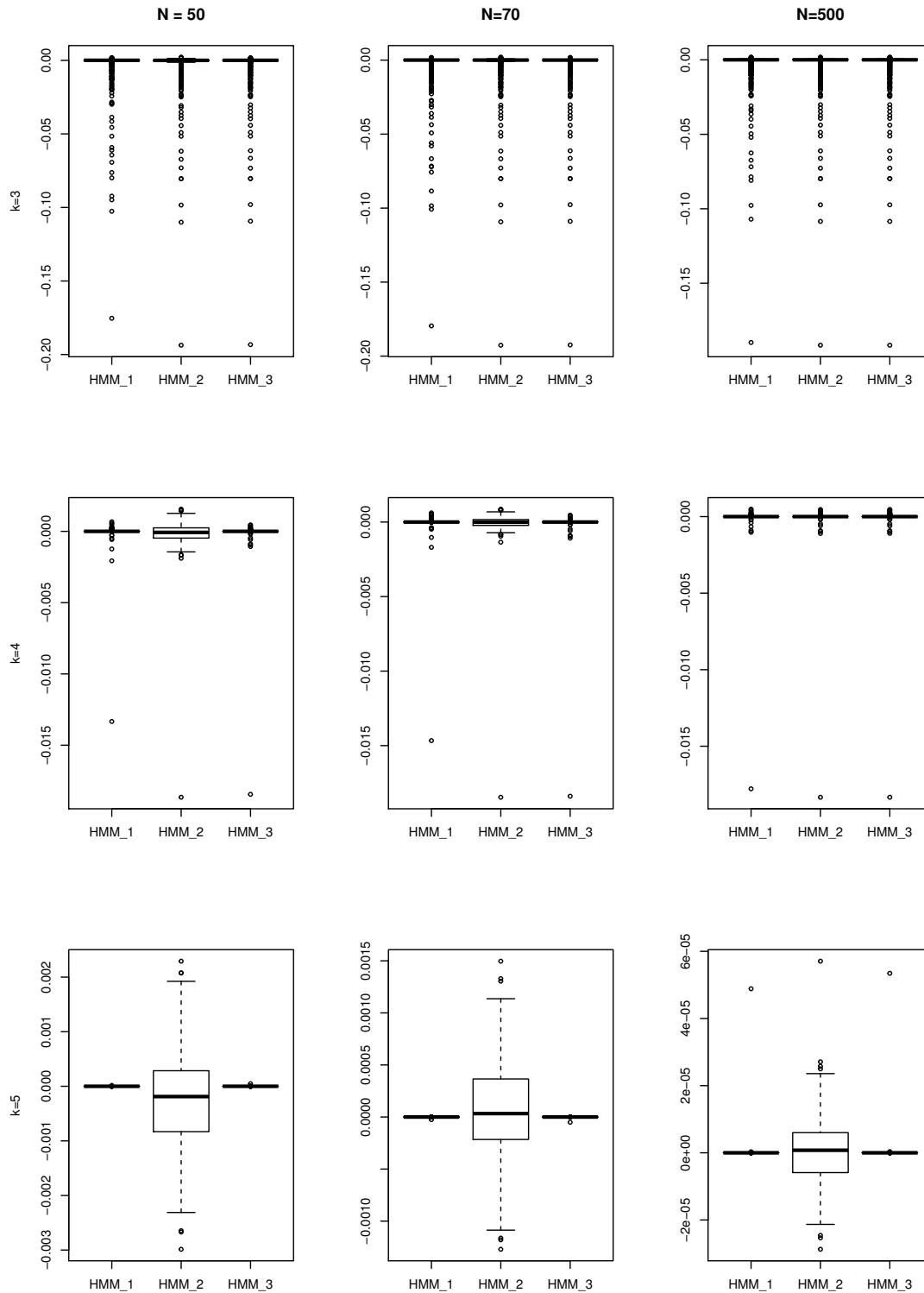


FIGURE 3.1. Box-plot des erreurs d'approximations de la log-vraisemblance du modèle $AR(1)$ bruité par *filtre HMM*, selon le nombre d'intervalles (N), le choix du nœud (HMM_1 , HMM_2 ou HMM_3) et des limites d'intégrations (k).

TABLEAU 3.3. Qualité d'approximation de la log-vraisemblance du modèle $AR(1)$ bruité par *filtre particulaire continu* (CPF) et *bootstrap filter* (PF) selon le nombre de particules (M).

<i>Filtre</i>	M	$\overline{\log\text{-vrais}}$	$100 \times \bar{\Delta}$	$100 \times \Delta_{sd}$	<i>Temps de calcul</i> (s)
<i>Kalman</i>		745.0415			0.011
<i>CPF</i>	500	746.3285	-0.1728	0.1247	0.368
	10^3	745.6199	-0.0776	0.0937	0.846
	10^4	745.1027	-0.0082	0.0294	1.809
	10^5	745.0426	-0.0002	0.0082	36.399
	5×10^5	745.0398	0.0003	0.0047	152.875
<i>PF</i>	500	746.3225	-0.1719	0.1230	0.201
	10^3	745.6211	-0.0777	0.0934	0.719
	10^4	745.0976	-0.0082	0.0296	0.807
	10^5	745.0429	-0.0001	0.0082	24.519
	5×10^5	745.0399	0.0002	0.0047	102.325

Note : Le temps de calcul représente le temps moyen (en seconde) mis par l'approche pour évaluer une fois la log-vraisemblance avec M particules et pour une série de longueur $T = 1000$.

Aussi, comparativement aux filtres HMM_1 et HMM_3 , le filtre HMM_2 (approximation de [Langrock et al. \(2012\)](#)) s'est révélé moins précis et nécessite plus de maillage (N) de l'espace d'états pour atteindre le même niveau de précision. La figure [3.1](#) offre un aperçu de la qualité de l'approximation en fonction du paramètre k et du nombre d'intervalles.

On ne note quasiment pas de différence entre les qualités d'approximation et les approximations de la log-vraisemblance par les deux filtres particuliers. Comparativement aux tableaux [3.1](#) et [3.2](#), il ressort du tableau [3.3](#) que le *filtre particulaire continu* et le *bootstrap filter* nécessitent un grand nombre de particules pour atteindre un niveau de précision comparable à celui du *filtre HMM*. À un niveau de précision égal, il est donc moins coûteux en temps d'approximer la log-vraisemblance par *filtre HMM* que par filtres particuliers, comme l'illustrent les temps de calcul.

Il est aussi important de souligner que l'approximation est d'autant plus précise que le nombre de particules est élevé. Cette propriété asymptotique de l'approche est cohérente

avec les résultats de convergence du filtre particulaire [Crişan et Doucet \(2002\)](#), [Del Moral et Miclo \(2000\)](#) et [Crişan et al. \(1999\)](#).

3.3. COMPARAISON DES APPROXIMATIONS DE LOG-VRAISEMBLANCE DES MODÈLES À VOLATILITÉ STOCHASTIQUE

Cette section offre une comparaison des approximations de la log-vraisemblance des modèles à volatilité stochastique selon les deux approches. On s'intéresse en particulier à la qualité de l'approximation de la log-vraisemblance de ces modèles par l'approche du *filtre HMM* comparativement au *filtre particulaire continu*. En effet, les propriétés asymptotiques du filtre particulaire nous assurent que pour un grand nombre de particules, l'approximation de la log-vraisemblance obtenue par cette approche peut être considérée comme «exacte».

3.3.1. Données et jeux de paramètres

Dans la présente étude, nous avons utilisé $M = 1 \times 10^5$ particules pour évaluer la log-vraisemblance des modèles SV_0 , SVL et $SVLJ$ par *filtre particulaire continu* et *bootstrap filter*. La moyenne de dix évaluations de la log-vraisemblance avec différentes particules a été utilisée comme approximation de la log-vraisemblance dans chaque cas. Les trois différents filtres HMM_1 , HMM_2 et HMM_3 ont été analysés avec les six valeurs du nombre d'intervalles discrets N et la valeur 5 pour le paramètre k ($k = 5$).

Pour l'approximation des log-vraisemblances, les trois jeux de paramètres suivants ont été considérés pour chaque modèle :

- SV_0 ($\alpha, \phi, \sigma_\eta^2$) :

(0.0125, 0.975, 0.02) ; (0.00625, 0.975, 0.025) ; (0.005, 0.98, 0.04)

- SVL ($\alpha, \phi, \sigma_\eta^2, \rho$) :

(0.0125, 0.975, 0.02, -0.8) ; (0.00625, 0.975, 0.025, -0.8) ; (0.005, 0.98, 0.04, -0.8)

- *SVLJ* $(\alpha, \phi, \sigma_\eta^2, \rho, p, \sigma_J^2)$:

$(0.0125, 0.975, 0.02, -0.8, 0.10, 10)$; $(0.00625, 0.975, 0.025, -0.8, 0.10, 0.5)$;

$(0.005, 0.98, 0.04, -0.8, 0.01, 10)$.

Ces valeurs sont inspirées des tableaux 2, 3 et 4 de [Pitt et al. \(2014\)](#) et sont en accord avec la littérature empirique en économétrie financière.

Nos analyses ont été réalisées sur les données des rendements en pourcentage de l'indice *S&P 500* sur la période allant du 3 janvier 2000 au 30 juin 2016 ($T = 4150$, source : *Federal Reserve Economic Data* (FRED)).

Les résultats de ces analyses sont résumés dans les tableaux [3.4](#), [3.5](#) et [3.6](#) ci-dessous, qui offrent une comparaison des approches par filtres HMM et par filtres particulières pour l'approximation de la log-vraisemblance des modèles SV_0 , SVL et SVLJ.

3.3.2. Résultats des études sur l'approximation de la log-vraisemblance des modèles SV_0 , SVL et SVLJ

Nous avons rapporté dans les tableaux [3.4](#), [3.5](#) et [3.6](#) les log-vraisemblances approximées et les erreurs relatives d'approximations ($\Delta = \frac{\hat{\ell}_{CPF} - \hat{\ell}_{HMM}}{|\hat{\ell}_{CPF}|}$) pour les modèles à volatilité stochastique sur la série S&P 500. En général, il en ressort que les deux approches génèrent pratiquement la même approximation de la log-vraisemblance de ces modèles.

En effet, pour les trois jeux de paramètres et les trois modèles à volatilité stochastique, quasiment toutes les erreurs relatives sont au plus de l'ordre de 10^{-3} . La seule exception reste toutefois l'écart considérable entre le résultat d'approximation du filtre HMM₃ comparativement aux autres filtres (HMM₁, HMM₂, *PF* et *CPF*) pour le modèle SVLJ (tableau [3.6](#)).

Les approximations par *filtre particulaire continu* et par *bootstrap filter* produisent des résultats similaires. Même si les temps de calculs n'ont pas été rapportés, nous avons généralement noté que les approches par filtres particulières nécessitent plus de temps de calculs.

De plus, le *bootstrap filter* est plus rapide que le *filtre particulaire continu* à cause du surcoût qu'engendre l'étape de rééchantillonnage continu .

TABLEAU 3.4. Comparaison de l'approximation de la log-vraisemblance du modèle SV_0 , selon les approches par filtres HMM et par filtres particulières, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM₁, HMM₂ ou HMM₃). Le paramètre k est fixé à 5.

<i>Filtre</i>	N	$\log\text{-vrais}$	$100 \times \Delta$	N	$\log\text{-vrais}$	$100 \times \Delta$
$\theta_1 = (0.0125, 0.975, 0.02, -0.8, 0.10, 10)$						
<i>CPF</i>		-5918.786		<i>PF</i>	-5918.789	
HMM ₁		-5918.646	-0.0024		-5918.788	0.0001
HMM ₂	30	-5909.523	-0.1565	100	-5917.772	-0.0171
HMM ₃		-5918.621	-0.0028		-5918.781	-0.0001
HMM ₁		-5918.795	0.0002		-5918.784	-0.0001
HMM ₂	50	-5914.975	-0.0644	300	-5918.667	-0.0020
HMM ₃		-5918.780	-0.0001		-5918.782	-0.0001
HMM ₁		-5918.791	0.0001		-5918.783	-0.0001
HMM ₂	70	-5916.764	-0.0342	500	-5918.740	-0.0008
HMM ₃		-5918.781	-0.0001		-5918.782	-0.0001
$\theta_2 = (0.00625, 0.975, 0.025, -0.8, 0.10, 0.5)$						
<i>CPF</i>		-5895.763		<i>PF</i>	-5895.759	
HMM ₁		-5895.633	-0.0022		-5895.712	-0.0009
HMM ₂	30	-5891.237	-0.0768	100	-5895.179	-0.0099
HMM ₃		-5895.618	-0.0025		-5895.709	-0.0009
HMM ₁		-5895.717	-0.0008		-5895.710	-0.0009
HMM ₂	50	-5893.757	-0.0340	300	-5895.649	-0.0019
HMM ₃		-5895.708	-0.0009		-5895.709	-0.0009
HMM ₁		-5895.714	-0.0008		-5895.709	-0.0009
HMM ₂	70	-5894.658	-0.0187	500	-5895.687	-0.0013
HMM ₃		-5895.708	-0.0009		-5895.709	-0.0009
$\theta_3 = (-0.003125, 0.975, 0.045, -0.8, 0.2, 0.5)$						
<i>CPF</i>		-5885.728		<i>PF</i>	-5885.713	
HMM ₁		-5885.629	-0.0017		-5885.742	0.0007
HMM ₂	30	-5889.126	0.0577	100	-5885.980	0.0043
HMM ₃		-5885.631	-0.0016		-5885.742	0.0002
HMM ₁		-5885.741	0.0002		-5885.742	0.0002
HMM ₂	50	-5886.783	0.0179	300	-5885.768	0.0007
HMM ₃		-5885.742	0.0002		-5885.742	0.0002
HMM ₁		-5885.741	0.0002		-5885.742	0.0002
HMM ₂	70	-5886.243	0.0088	500	-5885.751	0.0004
HMM ₃		-5885.742	0.0002		-5885.742	0.0002

Note : HMM₁, HMM₂ et HMM₃ correspondent respectivement aux approximations spécifiées par les expressions (2.3.5), (2.3.6) et (2.3.7) et $\Delta = \frac{\hat{\ell}_{CPF} - \hat{\ell}_{HMM}}{|\hat{\ell}_{CPF}|}$.

TABLEAU 3.5. Comparaison de l'approximation de la log-vraisemblance du modèle SVL, selon les approches par filtres HMM et par filtres particuliers, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM₁, HMM₂ ou HMM₃). Le paramètre k est fixé à 5.

<i>Filtre</i>	N	\log -vrais	$100 \times \Delta$	N	\log -vrais	$100 \times \Delta$
$\theta_1 = (0.0125, 0.975, 0.02, -0.8, 0.10, 10)$						
<i>CPF</i>		-5859.881		<i>PF</i>	-5859.836	
HMM ₁		-5859.628	-0.0043		-5859.855	-0.0004
HMM ₂	30	-5828.734	-0.5315	100	-5855.813	-0.0694
HMM ₃		-5859.628	-0.0043		-5859.855	-0.0004
HMM ₁		-5859.868	-0.0002		-5859.855	-0.0004
HMM ₂	50	-5845.516	-0.2451	300	-5859.389	-0.0084
HMM ₃		-5859.868	-0.0002		-5859.855	-0.0004
HMM ₁		-5859.855	-0.0004		-5859.855	-0.0004
HMM ₂	70	-5851.953	-0.1353	500	-5859.686	-0.0033
HMM ₃		-5859.855	-0.0004		-5859.855	-0.0004
$\theta_2 = (0.00625, 0.975, 0.025, -0.8, 0.10, 0.5)$						
<i>CPF</i>		-5804.610		<i>PF</i>	-5804.598	
HMM ₁		-5804.785	0.0030		-5804.621	0.0002
HMM ₂	30	-5790.753	-0.2387	100	-5802.651	-0.0337
HMM ₃		-5804.786	0.0030		-5804.621	0.0002
HMM ₁		-5804.637	0.0005		-5804.621	0.0002
HMM ₂	50	-5797.831	-0.1168	300	-5804.392	-0.0038
HMM ₃		-5804.637	0.0005		-5804.621	0.0002
HMM ₁		-5804.621	0.0002		-5804.621	0.0002
HMM ₂	70	-5800.807	-0.0655	500	-5804.538	-0.0012
HMM ₃		-5804.621	0.0002		-5804.621	0.0002
$\theta_3 = (-0.003125, 0.975, 0.045, -0.8, 0.2, 0.5)$						
<i>CPF</i>		-5768.957		<i>PF</i>	-5768.951	
HMM ₁		-5768.987	0.0005		-5768.947	-0.0002
HMM ₂	30	-5773.171	0.0731	100	-5769.056	0.0017
HMM ₃		-5768.994	0.0007		-5768.947	-0.0002
HMM ₁		-5768.958	0.0000		-5768.947	-0.0002
HMM ₂	50	-5769.853	0.0155	300	-5768.955	0.0000
HMM ₃		-5768.958	0.0000		-5768.947	-0.0002
HMM ₁		-5768.947	-0.0002		-5768.947	-0.0002
HMM ₂	70	-5769.259	0.0053	500	-5768.949	-0.0002
HMM ₃		-5768.947	-0.0002		-5768.947	-0.0002

Note : HMM₁, HMM₂ et HMM₃ correspondent respectivement aux approximations spécifiées par les expressions (2.3.5), (2.3.6) et (2.3.7) et $\Delta = \frac{\hat{\ell}_{CPF} - \hat{\ell}_{HMM}}{|\hat{\ell}_{CPF}|}$.

TABLEAU 3.6. Comparaison de l'approximation de la log-vraisemblance du modèle SVLJ, selon les approches par filtres HMM et par filtres particuliers, pour trois jeux de paramètres et en fonction du nombre d'intervalles (N) et du choix du nœud (HMM₁, HMM₂ ou HMM₃). Le paramètre k est fixé à 5.

<i>Filtre</i>	N	$\log -vrais$	$100 \times \Delta$	N	$\log -vrais$	$100 \times \Delta$
$\theta_1 = (0.0125, 0.975, 0.02, -0.8, 0.10, 10)$						
<i>CPF</i>		-5989.129		<i>PF</i>	-5989.103	
HMM ₁		-5996.622	0.1251		-5997.374	0.1377
HMM ₂	30	-5961.950	-0.4538	100	-5992.844	0.0620
HMM ₃		-9102.224	51.9791		-6499.153	8.5158
HMM ₁		-5997.020	0.1318		-5997.528	0.1417
HMM ₂	50	-5981.008	-0.1356	300	-5997.093	0.1330
HMM ₃		-7427.006	24.0081		-6116.065	2.1194
HMM ₁		-5997.219	0.1351		-5997.617	0.1426
HMM ₂	70	-5988.380	-0.0125	500	-5997.477	0.1394
HMM ₃		-6846.387	14.3136		-6062.833	1.2306
$\theta_2 = (0.00625, 0.975, 0.025, -0.8, 0.10, 0.5)$						
<i>CPF</i>		-5805.154		<i>PF</i>	-5805.146	
HMM ₁		-5805.753	0.0103		-5806.444	0.0222
HMM ₂	30	-5790.003	-0.2610	100	-5804.198	-0.0165
HMM ₃		-8741.777	50.5865		-6193.771	6.6943
HMM ₁		-5806.070	0.0158		-5806.716	0.0269
HMM ₂	50	-5798.396	-0.1164	300	-5806.451	0.0223
HMM ₃		-7043.386	21.3299		-5883.373	1.3474
HMM ₁		-5806.276	0.0193		-5806.773	0.0279
HMM ₂	70	-5801.952	-0.0552	500	-5806.676	0.0262
HMM ₃		-6502.485	12.0123		-5846.433	0.7111
$\theta_3 = (-0.003125, 0.975, 0.045, -0.8, 0.2, 0.5)$						
<i>CPF</i>		-5768.195		<i>PF</i>	-5768.264	
HMM ₁		-5759.799	-0.1456		-5759.730	-0.1467
HMM ₂	30	-5762.350	-0.1013	100	-5759.640	-0.1483
HMM ₃		-8554.391	48.3027		-6056.722	5.0020
HMM ₁		-5759.742	0.1465		-5759.610	0.1488
HMM ₂	50	-5759.939	-0.1431	300	-5759.593	-0.1491
HMM ₃		-6842.028	18.6165		-5804.113	0.6227
HMM ₁		-5759.732	-0.1467		-5759.610	-0.1488
HMM ₂	70	-5759.659	-0.1480	500	-5759.603	-0.1490
HMM ₃		-6334.091	9.8106		-5779.773	0.2007

Note : HMM₁, HMM₂ et HMM₃ correspondent respectivement aux approximations spécifiées par les expressions (2.3.5), (2.3.6) et (2.3.7) et $\Delta = \frac{\hat{\ell}_{CPF} - \hat{\ell}_{HMM}}{|\hat{\ell}_{CPF}|}$.

Chapitre 4

ÉVALUATION DE LA QUALITÉ DE L'ESTIMATION DES PARAMÈTRES DES MODÈLES À VOLATILITÉS STOCHASTIQUES

Pour faire suite au chapitre précédent, nous allons nous intéresser dans le présent chapitre à l'évaluation de la performance de ces deux approches pour l'estimation des paramètres des modèles à volatilité stochastique par optimisation de la log-vraisemblance approximée. Dans cette optique, deux études de simulations ont été réalisées.

Dans un premier temps, une réplique de l'étude de [Jacquier et al. \(1994\)](#) (ci-après JPR) a été menée afin d'évaluer et de comparer à d'autres approches, les résultats du *filtre HMM* et du *filtre particulaire continu* dans l'estimation des paramètres du modèle SV_0 .

Pour les modèles SVL et SVLJ, nous avons réalisé un second groupe d'études à travers la reproduction des tableaux 2, 3 et 4 de l'article de [Pitt et al. \(2014\)](#).

4.1. PERFORMANCES DES DEUX APPROCHES DANS L'ESTIMATION DES PARAMÈTRES DU MODÈLE SV_0

À l'instar de plusieurs études sur l'estimation des paramètres du modèle SV_0 , nous avons repris ici l'étude Monte Carlo utilisée par [Jacquier et al. \(1994\)](#) pour évaluer et comparer la performance de l'approche MCMC à celle des méthodes *MM* et *QML*. En effet, cette

expérience a souvent été reprise dans la littérature des modèles à volatilité stochastique ce qui facilite la comparaison des résultats des approches par *filtre HMM* et par *filtre particulière continu* aux méthodes MCMC de [Jacquier et al. \(1994\)](#), de maximum de vraisemblance de [Fridman et Harris \(1998\)](#) et de MCL de [Sandmann et Koopman \(1998\)](#).

4.1.1. Expérience de simulation JPR

Nous avons simulé 500 séries de longueur $T = 500$ chacune selon le modèle SV_0 (1.1.1) avec neuf différents jeux de paramètres $\theta_i = (\alpha, \phi, \sigma_\eta)$, $i = 1, \dots, 9$. Les valeurs du paramètre ϕ sont choisies entre 0.90, 0.95 et 0.98, dans l'optique de refléter la forte persistance qu'exhibe la volatilité de la plupart des séries financières. Ensuite, pour chaque valeur du paramètre ϕ , les paramètres α et σ_η sont choisis de sorte que les coefficients de variation modifiés $CV = \frac{\text{Var}(h_i)}{\mathbb{E}[h_i]^2} = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1$ soient égaux à 10, 1 ou 0.1. Les neuf jeux de paramètres sont alors donnés par :

- $CV = 10$:

$$\theta_1 = (-0.821, 0.90, 0.675); \quad \theta_2 = (-0.411, 0.95, 0.484); \quad \theta_3 = (-0.164, 0.98, 0.308)$$

- $CV = 1$:

$$\theta_4 = (-0.736, 0.90, 0.363); \quad \theta_5 = (-0.368, 0.95, 0.26); \quad \theta_6 = (-0.147, 0.98, 0.166)$$

- $CV = 0.1$:

$$\theta_7 = (-0.706, 0.90, 0.135); \quad \theta_8 = (-0.353, 0.95, 0.096); \quad \theta_9 = (-0.141, 0.98, 0.061)$$

Pour chacune des 500 séries, la log-vraisemblance a été approximée selon l'approche du *filtre HMM* d'une part et celle du *filtre particulière continu* d'autre part. L'approche du maximum de vraisemblance a été utilisée par la suite pour estimer les paramètres du modèle SV_0 selon chaque approche.

4.1.2. Présentation des résultats de l'étude JPR

Pour chaque jeu de paramètres, la moyenne et, entre parenthèse, la racine de l'erreur quadratique moyenne (*RMSE*) des paramètres estimés pour les 500 séries selon les deux approches sont rapportées dans le tableau 4.1. Ce tableau contient les résultats pour $N = 30$ nœuds et $M = 500$ particules; ces résultats sont similaires à ceux obtenus avec d'autres valeurs de ces deux paramètres. En plus des résultats des approches par *filtre HMM* et par *filtre particulaire continu*, le tableau 4.1 offre une comparaison aux résultats des méthodes MCMC, par maximum de vraisemblance (F&H) et MCL. Les entrées de ces dernières approches proviennent respectivement de la table 7 de [Jacquier et al. \(1994\)](#), de la table 1 de [Fridman et Harris \(1998\)](#) et de la table 2 de [Sandmann et Koopman \(1998\)](#).

D'une manière générale, le *filtre HMM* et le *filtre particulaire continu* présentent des performances semblables aux trois autres approches d'estimation. Les paramètres α et ϕ sont toujours bien estimés respectivement par la méthode MCL et par l'approche par *filtre HMM*. On note une détérioration de la précision des paramètres estimés avec la diminution du coefficient de variation.

Comparativement aux résultats de [Fridman et Harris \(1998\)](#), les résultats du *filtre HMM* sont légèrement plus précis. La méthodologie étant similaire, cette différence de précision entre les deux approches provient du choix de k dans la définition des limites extrêmes x_0 et x_N ($k = 5$ vs. $k = 3$) et du choix des nœuds pour chaque intervalle (points inférieurs vs. méthode Gauss-Legendre).

4.2. PERFORMANCE DES DEUX APPROCHES DANS L'ESTIMATION DES PARAMÈTRES DES MODÈLES SVL ET SVLJ

Nous avons repris l'expérience de simulation présentée dans la section précédente avec quelques modifications dans l'optique d'évaluer la performance du *filtre HMM* et du *filtre*

TABLEAU 4.1. Comparaison des approches MCMC, MCL, F&H, *filtre particulaire continu* (CPF) avec $M = 500$ particules et *filtre HMM* (HMM_1) avec $N = 30$ intervalles pour l'estimation des paramètres du modèle SV_0 .

	α	ϕ	σ_η	α	ϕ	σ_η	α	ϕ	σ_η
$CV = 10$									
<i>Vraie</i>	-0.821	0.900	0.675	-0.411	0.950	0.484	-0.164	0.980	0.308
MCL	-0.837 (0.034)	0.915 (0.025)	0.579 (0.119)	-0.417 (0.021)	0.953 (0.020)	0.436 (0.077)	-0.166 (0.010)	0.977 (0.020)	0.290 (0.053)
MCMC	-0.679 (0.220)	0.916 (0.026)	0.562 (0.120)	-0.464 (0.160)	0.940 (0.020)	0.460 (0.055)	-0.190 (0.080)	0.980 (0.010)	0.350 (0.060)
F&H	-0.896 (0.280)	0.890 (0.034)	0.685 (0.080)	-0.505 (0.180)	0.940 (0.020)	0.495 (0.070)	-0.100 (0.080)	0.986 (0.010)	0.320 (0.050)
HMM_1	-0.880 (0.272)	0.893 (0.033)	0.678 (0.084)	-0.422 (0.169)	0.948 (0.020)	0.489 (0.070)	-0.231 (0.135)	0.972 (0.016)	0.316 (0.053)
CPF	-0.908 (0.287)	0.889 (0.034)	0.676 (0.084)	-0.497 (0.215)	0.940 (0.026)	0.489 (0.070)	-0.250 (0.164)	0.970 (0.019)	0.316 (0.053)
$CV = 1$									
<i>Vraie</i>	-0.736	0.900	0.363	-0.368	0.950	0.26	-0.147	0.980	0.166
MCL	-0.745 (0.022)	0.897 (0.100)	0.325 (0.080)	-0.372 (0.011)	0.93 (0.102)	0.233 (0.075)	-0.148 (0.010)	0.97 (0.071)	0.161 (0.050)
MCMC	-0.870 (0.340)	0.880 (0.046)	0.350 (0.067)	-0.560 (0.340)	0.920 (0.046)	0.280 (0.065)	-0.220 (0.140)	0.970 (0.020)	0.230 (0.080)
F&H	-0.870 (0.430)	0.880 (0.050)	0.370 (0.080)	-0.510 (0.306)	0.930 (0.040)	0.280 (0.070)	-0.090 (0.060)	0.987 (0.015)	0.180 (0.040)
HMM_1	-0.741 (0.446)	0.899 (0.059)	0.377 (0.093)	-0.427 (0.343)	0.942 (0.044)	0.273 (0.069)	-0.124 (0.086)	0.983 (0.012)	0.179 (0.049)
CPF	-0.879 (0.515)	0.881 (0.068)	0.372 (0.088)	-0.491 (0.388)	0.933 (0.050)	0.272 (0.067)	-0.244 (0.186)	0.967 (0.025)	0.177 (0.048)
$CV = .1$									
<i>Vraie</i>	-0.706	0.900	0.135	-0.353	0.95	0.096	-0.141	0.980	0.061
MCL	-0.709 (0.010)	0.443 (0.770)	0.156 (0.112)	-0.355 (0.010)	0.526 (0.735)	0.136 (0.108)	-0.142 (0.001)	0.572 (0.726)	0.113 (0.113)
MCMC	-1.540 (1.350)	0.780 (0.190)	0.150 (0.082)	-1.120 (1.150)	0.840 (0.160)	0.120 (0.074)	-0.660 (0.830)	0.910 (0.120)	0.140 (0.099)
F&H	-1.360 (1.720)	0.810 (0.240)	0.160 (0.120)	-0.810 (1.150)	0.886 (0.160)	0.120 (0.090)	-0.537 (1.130)	0.924 (0.160)	0.088 (0.090)
HMM_1	-0.653 (1.406)	0.908 (0.197)	0.095 (0.115)	-0.341 (0.936)	0.952 (0.132)	0.056 (0.084)	-0.263 (0.828)	0.963 (0.116)	0.059 (0.072)
CPF	-1.592 (1.953)	0.775 (0.323)	0.158 (0.108)	-0.812 (1.112)	0.891 (0.117)	0.106 (0.088)	-0.663 (0.913)	0.922 (0.109)	0.055 (0.069)

Note : Les moyennes et les *RMSE* (entre parenthèses) des paramètres estimés sont présentés. Les meilleurs estimés sont en gras. Les entrées MCMC, F&H et MCL proviennent respectivement de table 7 [Jacquier et al. \(1994\)](#), table 1 [Fridman et Harris \(1998\)](#) et de table 2 [Sandmann et Koopman \(1998\)](#).

particulaire continu dans le cadre de l'estimation des paramètres des modèles à volatilité stochastique avec effet de levier (SVL) et avec sauts (SVLJ).

4.2.1. Expérience de simulation

Pour pouvoir comparer nos résultats dans le cas du modèle SVLJ, nous allons utiliser les mêmes paramètres que ceux des tables 2, 3 et 4 de Pitt et al. (2014). À la place du paramètre α , les auteurs ont plutôt utilisé le paramètre μ tel que $\alpha = \mu \times (1 - \phi)$. L'expérience a été menée avec deux jeux de paramètres $\theta_i = (\mu, \phi, \sigma_\eta^2, \rho)$, $i = 1, 2$ pour le modèle SVL et trois $\theta_i = (\mu, \phi, \sigma_\eta^2, \rho, p, \sigma_f^2)$, $i = 1, 2, 3$ pour le modèle SVLJ :

- SVL : $(0.5, 0.975, 0.02, -0.8)$; $(0.25, 0.975, 0.025, -0.8)$

- SVLJ : $(0.5, 0.975, 0.02, -0.8, 0.10, 10)$; $(0.25, 0.975, 0.025, -0.8, 0.10, 0.5)$;
 $(0.25, 0.975, 0.025, -0.8, 0.01, 10)$.

Pour chaque jeu de paramètres, nous avons simulé 50 séries de longueur $T = 2000$ selon le modèle SVL (1.1.1) d'une part et le modèle SVLJ (1.2.1) d'autre part.

Nous avons par la suite, comme dans la section précédente approximé la log-vraisemblance de chacune des 50 séries par le *filtre HMM* et le *filtre particulaire continu*. L'estimation des paramètres des modèles SVL et SVLJ a été réalisée par optimisation des log-vraisemblances approximées selon chaque approche.

4.2.2. Présentation des résultats de l'expérience

Les résultats de ces simulations sont résumés dans le tableau 4.2 pour le modèle SVL et le tableau 4.3 pour le modèle SVLJ. Nous avons présenté les résultats pour $N = 30$ nœuds et pour $M = 500$ particules. Dans chaque cas, la moyenne et, entre parenthèses, la racine de l'erreur quadratique moyenne (*RMSE*) des paramètres estimés pour les 50 séries sont

rapportées. Les résultats des tables 2, 3 et 4 de Pitt et al. (2014) sont aussi rapportés en guise de comparaison dans le tableau 4.3. Les *RMSE* relatives aux résultats Pitt et al. (2014) ont été calculées à partir du biais et de la variance des erreurs des tables 2, 3 et 4.

TABLEAU 4.2. Résultats de l'estimation des paramètres du modèle SVL par *filtre particulaire continu* (CPF) avec $M = 500$ particules et *filtre HMM* (HMM_1) avec $N = 30$ intervalles sur les données simulées pour deux jeux de paramètres.

	μ	ϕ	σ_η^2	ρ
<i>Vraie</i>	0.5	0.975	0.02	-0.8
HMM_1	0.514 (0.09)	0.972 (0.006)	0.021 (0.005)	-0.789 (0.055)
CPF	0.52 (0.092)	0.974 (0.005)	0.020 (0.004)	-0.799 (0.059)
	μ	ϕ	σ_η^2	ρ
<i>Vraie</i>	0.25	0.975	0.025	-0.8
HMM_1	0.267 (0.099)	0.972 (0.006)	0.026 (0.005)	-0.786 (0.05)
CPF	0.273 (0.102)	0.974 (0.005)	0.025 (0.005)	-0.8 (0.05)

Note : Moyennes et entre parenthèses les *RMSE* des paramètres estimés sur 50 séries de longueur $T = 2000$ sont présentés.

Comme l'illustre le tableau 4.2, les deux approches présentent des performances similaires pour l'estimation des paramètres du modèle SVL. Les biais d'estimation sont négligeables pour les deux jeux de paramètres considérés. Le paramètre μ génère les plus grandes *RMSE* pour les deux approches. Les conclusions sont similaires sur la performance des méthodes dans l'estimation des paramètres du modèle SVLJ (4.3). Nos résultats par l'approche du *filtre particulaire continu* sont relativement meilleurs que ceux rapportés par Pitt et al. (2014). Comparativement aux résultats des approches par *filtre particulaire continu*, on note une meilleure précision de l'approche par *filtre HMM* pour l'estimation des paramètres ρ et σ_J^2 .

TABLEAU 4.3. Résultats de l'estimation des paramètres du modèle SVLJ par *filtre particulaire continu* (CPF) avec $M = 500$ particules et *filtre HMM* (HMM_1) avec $N = 30$ intervalles sur les données simulées pour trois différents jeux de paramètres.

	μ	ϕ	σ_η^2	ρ	p	σ_J^2
<i>Vraie</i>	0.5	0.975	0.02	-0.8	0.1	10
HMM_1	0.515 (0.105)	0.973 (0.01)	0.017 (0.007)	-0.752 (0.259)	0.102 (0.023)	9.968 (1.989)
CPF	0.558 (0.118)	0.974 (0.002)	0.019 (0.003)	-0.802 (0.013)	0.101 (0.010)	9.902 (1.448)
<i>Pitt et al., 2014 (Table 2)</i>	0.500 (0.125)	0.972 (0.011)	0.022 (0.008)	-0.837 (1.641)	0.104 (0.025)	9.801 (2.533)
	μ	ϕ	σ_η^2	ρ	p	σ_J^2
<i>Vraie</i>	0.25	0.975	0.025	-0.8	0.1	0.5
HMM_1	0.216 (0.182)	0.974 (0.005)	0.025 (0.006)	-0.800 (0.076)	0.302 (0.472)	0.526 (0.317)
CPF	0.223 (0.482)	0.977 (0.008)	0.028 (0.009)	-0.818 (1.041)	0.121 (0.214)	0.589 (0.374)
<i>Pitt et al., 2014 (Table 3)</i>	0.212 (0.169)	0.973 (0.009)	0.029 (0.013)	-0.856 (1.658)	0.235 (0.294)	0.633 (0.787)
	μ	ϕ	σ_η^2	ρ	p	σ_J^2
<i>Vraie</i>	0.25	0.975	0.025	-0.8	0.01	10
HMM_1	0.266 (0.110)	0.974 (0.005)	0.022 (0.006)	-0.781 (0.084)	0.024 (0.049)	10.625 (6.211)
CPF	0.257 (0.122)	0.974 (0.006)	0.027 (0.009)	-0.788 (0.914)	0.017 (0.019)	9.875 (6.041)
<i>Pitt et al., 2014 (Table 4)</i>	0.254 (0.118)	0.973 (0.009)	0.027 (0.009)	-0.823 (1.624)	0.013 (0.015)	9.620 (6.244)

Note : Moyennes et entre parenthèses les *RMSE* des paramètres estimés sur 50 séries de longueur $T = 2000$ sont présentés. Les entrées CPF (Table 2), CPF (Table 3) et CPF (Table 4) sont obtenues à partir des tables 2, 3 et 4 de [Pitt et al. \(2014\)](#).

Chapitre 5

ESTIMATIONS SUR DES DONNÉES EMPIRIQUES

Dans ce dernier chapitre nous allons exploiter les approches par *filtre HMM* et par *filtre particulière continu* pour estimer les paramètres des modèles à volatilité stochastique sur les données de rendements en pourcentage des indices S&P 500 et NASDAQ et du taux de change USD/EUR. Une comparaison de la qualité d'adéquation aux données des modèles à volatilité stochastique par rapport aux modèles GARCH et *MS-GARCH* (GARCH avec changement de régimes) est aussi présentée.

5.1. ESTIMATION DES PARAMÈTRES

La période allant du 3 janvier 2000 au 30 juin 2016 a été considérée pour les trois séries (source : *Federal Reserve Economic Data* (FRED)). Les trois modèles à volatilité stochastique SV_0 , SVL et $SVLJ$ ont été estimés sur chacune des séries. Dans chaque cas, nous avons rapporté les paramètres estimés, leurs erreurs types (entre parenthèses) ainsi que la valeur de la log-vraisemblance (*log-vrais*) en ces paramètres. Les critères AIC (*Akaike Information Criterion*) et BIC (*Bayesian Information Criterion*) ont aussi été rapportés pour la comparaison des modèles et l'appréciation de l'apport de l'effet de levier et du saut. Les valeurs rapportées dans les tableaux correspondent à la moitié de la valeur de ces critères.

Nous avons utilisé $N = 30$ et $N = 100$ nœuds pour l'estimation des paramètres des différents modèles par l'approche du *filtre HMM* et $M = 500$ et $M = 1000$ en ce qui concerne celle par *filtre particulaire continu*.

TABLEAU 5.1. Résultats de l'estimation des paramètres des modèles à volatilité stochastique par *filtre particulaire continu* (CPF) et *filtre HMM* sur la série S&P 500 pour la période du 2000-01-03 au 2016-06-30.

HMM_1	N	μ	ϕ	σ_η^2	ρ	p	σ_J^2	log-vrais	AIC	BIC	$\ell_{HMM}(500)$
SV_0	30	-0.078 (0.174)	0.984 (0.003)	0.030 (0.025)	- -	- -	- -	-5881.397 -	5884.397	5893.893	-5881.697
	100	-0.079 (0.169)	0.985 (0.004)	0.028 (0.029)	- -	- -	- -	-5881.567 -	5884.567	5894.063	-5881.556
SVL	30	-0.134 (0.084)	0.974 (0.003)	0.048 (0.028)	-0.798 (0.023)	- -	- -	-5768.845 -	5772.845	5785.507	-5769.026
	100	-0.125 (0.085)	0.976 (0.003)	0.045 (0.030)	-0.823 (0.027)	- -	- -	-5768.661 -	5772.661	5785.323	-5768.654
$SVLJ$	30	-0.818 (0.214)	0.977 (0.006)	0.063 (0.053)	-0.822 (0.025)	0.541 (0.088)	0.382 (0.082)	-5751.691 -	5757.691	5776.684	-5751.695
	100	-1.345 (0.470)	0.985 (0.004)	0.067 (0.032)	-0.869 (0.029)	0.605 (0.067)	0.399 (0.066)	-5749.607 -	5755.607	5774.600	-5749.423
CPF	M	μ	ϕ	σ_η^2	ρ	p	σ_J^2	log-vrais	AIC	BIC	$\ell_{CPF}(10^5)$
SV_0	500	-0.109 (0.118)	0.986 (0.002)	0.027 (0.012)	- -	- -	- -	-5880.089 -	5883.089	5892.585	-5881.595
	10^3	-0.050 (0.140)	0.984 (0.002)	0.030 (0.014)	- -	- -	- -	-5881.462 -	5884.462	5893.958	-5881.627
SVL	500	-0.141 (0.075)	0.977 (0.002)	0.045 (0.017)	-0.817 (0.017)	- -	- -	-5768.705 -	5772.705	5785.367	-5768.766
	10^3	-0.113 (0.162)	0.976 (0.002)	0.046 (0.048)	-0.819 (0.025)	- -	- -	-5770.520 -	5774.520	5787.182	5768.707
$SVLJ$	500	-0.125 (0.087)	0.979 (0.002)	0.050 (0.021)	-0.905 (0.012)	0.275 (0.014)	0.378 (0.052)	-5757.122 -	5763.122	5782.115	-5750.218
	10^3	-0.123 (0.078)	0.980 (0.001)	0.051 (0.020)	-0.885 (0.003)	0.191 (0.016)	0.423 (0.095)	-5761.272 -	5767.272	5786.265	-5762.823

Note : Moyennes et entre parenthèses les erreurs types des paramètres estimés. Ces derniers sont obtenus à partir de la matrice hessienne retournée par la routine d'optimisation. Les valeurs optimales sont en gras où $\ell_{HMM}(500)$ est la log-vraisemblance selon le *filtre HMM* avec $N = 500$ nœuds et $\ell_{CPF}(10^5)$ celle selon le *filtre particulaire* avec $M = 10^5$ particules.

TABLEAU 5.2. Résultats de l'estimation des paramètres des modèles à volatilité stochastique par *filtre particulaire continu* (CPF) et *filtre HMM* sur la série NASDAQ pour la période du 2000-01-03 au 2016-06-30.

HMM_1	N	μ	ϕ	σ_η^2	ρ	p	σ_J^2	log-vrais	AIC	BIC	$\ell_{HMM}(500)$
SV_0	30	0.662 (0.308)	0.990 (0.001)	0.025 (0.032)	-	-	-	-7266.905 -	7269.905	7279.401	-7270.657
	100	0.655 (0.285)	0.993 (0.002)	0.014 (0.025)	-	-	-	-7267.586 -	7270.586	7280.082	-7267.984
SVL	30	0.639 (0.107)	0.981 (0.002)	0.032 (0.024)	-0.725 (0.031)	-	-	-7195.268 -	7199.268	7211.929	-7193.451
	100	0.686 (0.117)	0.986 (0.002)	0.026 (0.027)	-0.774 (0.037)	-	-	-7191.537 -	7195.537	7208.198	-7191.556
$SVLJ$	30	0.225 (0.178)	0.983 (0.002)	0.038 (0.032)	-0.711 (0.035)	0.492 (0.109)	0.704 (0.109)	-7176.666 -	7182.666	7201.658	-7174.715
	100	-0.092 (0.313)	0.989 (0.002)	0.040 (0.034)	-0.786 (0.039)	0.635 (0.071)	0.693 (0.076)	-7168.945 -	7174.945	7193.937	-7175.948
CPF	M	μ	ϕ	σ_η^2	ρ	p	σ_J^2	log-vrais	AIC	BIC	$\ell_{CPF}(10^5)$
SV_0	500	0.639 (0.192)	0.994 (0.002)	0.016 (0.029)	-	-	-	-7268.523 -	7271.523	7281.019	-7267.683
	10^3	1.029 (0.106)	0.993 (0.001)	0.004 (0.006)	-	-	-	-7269.163 -	7272.163	7281.659	-7304.581
SVL	500	0.673 (0.104)	0.987 (0.001)	0.025 (0.012)	-0.777 (0.022)	-	-	-7191.604 -	7195.604	7208.265	-7191.869
	10^3	0.991 (0.053)	0.986 (0.001)	0.007 (0.015)	-0.825 (0.050)	-	-	-7192.372 -	7196.372	7209.033	-7288.034
$SVLJ$	500	0.609 (0.129)	0.987 (0.002)	0.034 (0.022)	-0.902 (0.009)	0.455 (0.032)	0.682 (0.068)	-7173.194 -	7179.194	7198.186	-7175.973
	10^3	0.741 (0.073)	0.987 (0.001)	0.016 (0.004)	-0.841 (0.019)	0.695 (0.073)	0.044 (0.078)	-7174.790 -	7180.790	7199.782	-7202.233

Note : Moyennes et entre parenthèses les erreurs types des paramètres estimés. Ces derniers sont obtenus à partir de la matrice hessienne retournée par la routine d'optimisation. Les valeurs optimales sont en gras où $\ell_{HMM}(500)$ est la log-vraisemblance selon le *filtre HMM* avec $N = 500$ nœuds et $\ell_{CPF}(10^5)$ celle selon le *filtre particulaire* avec $M = 10^5$ particules.

L'évaluation de la log-vraisemblance à l'optimum pour $N = 500$ et pour $M = 10^5$ respectivement selon l'approche par *filtre HMM* ($\ell_{HMM}(500)$) et par *filtre particulaire continu* ($\ell_{CPF}(10^5)$) ont aussi été rapportées.

TABLEAU 5.3. Résultats de l'estimation des paramètres des modèles à volatilité stochastique par *filtre particulaire continu* (CPF) et *filtre HMM* (HMM₁) sur la série USD/EUR pour la période du 2000-01-03 au 2016-06-30.

<i>HMM</i> ₁	<i>N</i>	μ	ϕ	σ_η^2	ρ	<i>p</i>	σ_J^2	log-vrais	AIC	BIC	$\ell_{HMM}(500)$
<i>SV</i> ₀	30	-1.039 (0.171)	0.990 (0.001)	0.007 (0.018)	-	-	-	-3761.942 -	3764.942	3774.437	-3764.430
	100	-1.035 (0.169)	0.994 (0.002)	0.004 (0.017)	-	-	-	-3762.001 -	3765.001	3774.496	-3762.035
<i>SVL</i>	30	-1.032 (0.194)	0.991 (0.002)	0.007 (0.026)	-0.296 (0.158)	-	-	-3760.150 -	3764.150	3776.810	-3764.955
	100	-1.028 (0.163)	0.994 (0.002)	0.003 (0.017)	-0.169 (0.105)	-	-	-3760.696 -	3764.696	3777.356	-3761.760
<i>SVLJ</i>	30	-1.099 (0.091)	0.995 (0.0004)	0.021 (0.021)	-0.505 (0.045)	0.209 (0.102)	0.288 (0.156)	-3740.840 -	3746.84	3765.830	-3790.716
	100	-1.207 (0.208)	0.995 (0.002)	0.004 (0.018)	-0.185 (0.105)	0.149 (0.103)	0.349 (0.250)	-3747.386 -	3753.386	3772.376	-3747.443
<i>CPF</i>	<i>M</i>	μ	ϕ	σ_η^2	ρ	<i>p</i>	σ_J^2	log-vrais	AIC	BIC	$\ell_{CPF}(10^5)$
<i>SV</i> ₀	500	-1.062 (0.166)	0.995 (0.001)	0.004 (0.010)	-	-	-	-3760.229 -	3763.229	3772.724	-3762.153
	10 ³	0.187 (0.091)	0.995 (0.005)	0.001 (0.013)	-	-	-	-3761.292 -	3764.292	3773.787	-3863.227
<i>SVL</i>	500	-1.062 (0.039)	0.995 (0.006)	0.004 (0.006)	-0.149 (0.276)	-	-	-3759.079 -	3763.079	3775.739	-3760.898
	10 ³	0.183 (0.073)	0.995 (0.001)	0.001 (0.004)	-0.278 (0.095)	-	-	-3759.929 -	3763.929	3776.589	-3865.961
<i>SVLJ</i>	500	-1.194 (0.216)	0.995 (0.001)	0.004 (0.011)	-0.175 (0.104)	0.150 (0.020)	0.340 (0.101)	-3745.467 -	3752.467	3771.457	-3748.116
	10 ³	0.125 (0.080)	0.995 (0.001)	0.001 (0.004)	-0.275 (0.170)	0.053 (0.011)	0.0001 (0.294)	-3746.941 -	3752.941	3771.931	-3857.858

Note : Moyennes et entre parenthèses les erreurs types des paramètres estimés. Ces derniers sont obtenus à partir de la matrice hessienne retournée par la routine d'optimisation. Les valeurs optimales sont en gras où $\ell_{HMM}(500)$ est la log-vraisemblance selon le *filtre HMM* avec $N = 500$ nœuds et $\ell_{CPF}(10^5)$ celle selon le *filtre particulaire* avec $M = 10^5$ particules.

Les tableaux 5.1, 5.2 et 5.3 présentent les résultats de l'estimation des paramètres des différents modèles selon chacune des deux approches respectivement pour le S&P 500, le

NASDAQ et le USD/EUR. De manière générale, les log-vraisemblances aux paramètres estimés sont comparables pour les deux approches avec une légère supériorité de celles obtenues par *filtre HMM*. En ce qui concerne les paramètres estimés, à l'exception du paramètre μ pour lequel on note une certaine différence, les deux approches produisent des résultats semblables sur les différentes séries. En effet, comme dans le chapitre précédent, les *RMSE* sont plus élevées pour le paramètre μ . Sur la base des log-vraisemblances (*log-vrais*) et des critères AIC et BIC, l'approche par *filtre HMM* avec $N = 100$ nœuds est préférable aux autres pour les séries du S&P 500 et du NASDAQ. Pour la série du taux de change USD/EUR le meilleur résultat est obtenu avec la méthode par *filtre HMM* et $N = 30$ nœuds. Pour tous ces critères, l'utilisation de $M = 500$ particules pour l'estimation des paramètres selon l'approche par *filtre particulaire continu* est préférable à $M = 1000$ particules.

En terme d'apport de l'effet de levier sur la qualité d'ajustement du modèle aux données, les résultats sont en accord avec la littérature en économétrie financière. En effet, l'inclusion de l'effet de levier, d'une part, n'affecte pratiquement pas la qualité d'ajustement du modèle sur les données des rendements du taux de change USD/EUR ($AIC_{SVL(100)} - AIC_{SV_0(100)} = -0.305$), et d'autre part, l'améliore considérablement pour les données des rendements des indices S&P 500 ($AIC_{SVL(100)} - AIC_{SV_0(100)} = -111.906$) et NASDAQ ($AIC_{SVL(100)} - AIC_{SV_0(100)} = -71.884$). Concernant le saut, son inclusion est bénéfique pour toutes les séries. Les conclusions sont identiques pour les estimations selon l'approche par *filtre particulaire continu*.

Dans la prochaine section, nous offrons une comparaison de la qualité d'adéquation aux données de ces modèles aux modèles de type GARCH avec et sans changement de régimes.

5.2. QUALITÉ D'ADÉQUATION AUX DONNÉES

Nous allons dans cette dernière section présenter les résultats de l'étude sur la qualité d'adéquation du modèle SVLJ aux trois séries financières (S&P 500, NASDAQ et USD/EUR)

comparativement aux modèles : GJR-GARCH(1,1) (GJR- t) (Glosten et al. (1993)) et GJR-GARCH(1,1) à changement de régimes à deux états (MS-GJR- t). L'extension "- t " au nom du modèle vise à signifier que la distribution de Student- t a été adoptée pour l'innovation du modèle. Les résultats sont présentés dans le tableau 5.4 où les entrées GJR- t et MS-GJR- t proviennent de la table 1 de Augustyniak et al. (2017).

Pour toutes les trois séries, le tableau 5.4 montre que sur la base des critères AIC et BIC, les résultats du modèle SVLJ par l'approche du *filtre HMM* sont meilleurs que les autres. De plus, indépendamment de l'approche d'estimation adoptée, le modèle SVLJ offre toujours une meilleure adéquation aux données que les modèles GJR- t et MS-GJR- t .

TABLEAU 5.4. Comparaison de la qualité d'adéquation de différentes modélisations avec effet de levier de la volatilité des trois séries financières.

Modèles	<i>SVLJ(HMM)</i>	<i>SVLJ(PF)</i>	<i>GJR-t</i>	<i>MS-GJR-t</i>
N_p	6	6	5	11
S&P 500				
<i>log-vrais</i>	-5749.6	-5757.1	-5782.0	-5770.5
AIC	5755.6	5763.1	5787.0	5781.5
BIC	5774.6	5782.1	5887.1	5816.3
NASDAQ				
<i>log-vrais</i>	-7168.9	-7173.2	-7197.5	-7175.5
AIC	7174.9	7179.2	7202.5	7186.5
BIC	7193.9	7198.2	7218.3	7221.3
USD/EUR				
<i>log-vrais</i>	-3740.8	-3745.5	-3745.6	-3740.5
AIC	3746.8	3751.5	3750.6	3751.5
BIC	3765.8	3770.5	3766.4	3786.3

Note : N_p : Nombre de paramètres libres du modèle, *log-vrais* : maximum de la log-vraisemblance. Les entrées *GJR-t* et *MS-GJR-t* proviennent de la table 1 de Augustyniak et al. (2017). Les valeurs optimales sont en gras.

CONCLUSION

Bien qu'il soit plus flexible et mieux adapté à la modélisation de la volatilité des séries financières, le modèle à volatilité stochastique est très peu utilisé par les praticiens du domaine comparativement aux modèles de type GARCH. La raison principale de cette situation réside dans la difficulté d'estimation des modèles à volatilité stochastique par l'approche classique de maximisation de la vraisemblance. En effet, il est impossible de calculer analytiquement la vraisemblance qui est une intégrale multidimensionnelle (1.3.2) pour ce type de modèle. L'analyse et la proposition de solutions à ce problème a été le fil conducteur de ce mémoire.

En réalité, même si dans la littérature financière un certain nombre de travaux se sont penchés sur l'estimation des modèles à volatilité stochastique par maximum de vraisemblance en se basant, d'une part sur la méthodologie d'intégration numérique récursive introduite par [Kitagawa \(1987\)](#) ([Fridman et Harris \(1998\)](#); [Bartolucci et De Luca \(2001\)](#); [Clements et al. \(2004\)](#); [Langrock et al. \(2012\)](#)), et d'autre part sur la technique de simulation Monte Carlo séquentielle ou filtre particulaire ([Pitt et al. \(2014\)](#)), aucune étude approfondie et comparative de la performance de ces approches n'avait encore été menée.

La première contribution de ce mémoire a été d'unifier et de placer les approches déterministes d'intégration numérique de [Fridman et Harris \(1998\)](#); [Bartolucci et De Luca \(2001\)](#); [Clements et al. \(2004\)](#); [Langrock et al. \(2012\)](#) dans le cadre plus général de résolution des problèmes de filtrage non-linéaires à bruits non-gaussiens. Ceci nous a conduit à développer

et formuler l'algorithme du *filtre HMM* (2.3.1) qui est une approximation du filtre optimal dans le cas d'un modèle espace-état à variable d'état continue. Les études Monte Carlo ont révélé que ce *filtre HMM* est très rapide et permet d'approximer avec une très grande précision la log-vraisemblance des modèles à volatilité stochastique.

Notre seconde contribution à cette littérature a consisté à généraliser cette approche aux modèles à volatilité stochastique incorporant l'effet de levier (SVL) et les sauts (SVLJ). L'évaluation de la qualité de l'approche dans l'estimation des paramètres de ces différents modèles à volatilité stochastique a été étudiée à travers plusieurs études Monte Carlo et empiriques. Pour le modèle SV_0 nous avons reproduit l'étude de simulation de [Jacquier et al. \(1994\)](#) et comparé aussi bien la performance de l'approche par *filtre HMM* que l'approche par *filtre particulière continu* aux méthodes MCMC de [Jacquier et al. \(1994\)](#), de maximum de vraisemblance de [Fridman et Harris \(1998\)](#) et de MCL de [Sandmann et Koopman \(1998\)](#). Nos résultats ont révélé que ces deux approches présentent des performances semblables aux trois autres approches d'estimation.

La troisième contribution de ce mémoire réside dans la comparaison des deux approches dans l'estimation des paramètres du modèle à volatilité stochastique avec effet de levier et sauts SVLJ. Dans cette optique, certaines études de l'article de [Pitt et al. \(2014\)](#) ont été reproduites et nous ont permis de noter que le *filtre HMM* présente des résultats similaires au *filtre particulière continu* pour l'estimation des paramètres du modèle SVLJ. En dernier lieu, nous avons présenté une étude comparative de la qualité d'adéquation aux données des modèles à volatilité stochastique par rapport aux modèles GARCH et GARCH avec changement de régimes (MS-GARCH).

Un premier avantage de l'approche par *filtre HMM* est le considérable gain en temps de calcul comparativement aux approches stochastiques qui sont très exigeantes en puissance et temps de calcul (MCMC, MCL, CPF). Au-delà de cet aspect, cette approche offre

aux praticiens un moyen efficace et simple d'estimer les paramètres des modèles à volatilité stochastique. En plus du fait qu'aucun effort supplémentaire n'est requis pour adopter d'autres distributions pour les innovations du système, le lien de l'approche avec les *HMM* assure l'obtention, sans grande difficulté, de la prévision et du filtrage des états. Par exemple, l'*algorithme de Viterbi* peut être utilisé pour déterminer la séquence optimale du processus latent $\{h_t\}$ au vu des observations $y_{1:t}$. Le gradient et la matrice hessienne de la fonction de log-vraisemblance du modèle peuvent aussi être dérivés récursivement et de manière exacte lors du calcul de la log-vraisemblance par l'algorithme du *filtre HMM* (2.3.1) en exploitant les résultats de [Lystig et Hughes \(2002\)](#).

Enfin, l'approche d'estimation des modèles à volatilité stochastique par *filtre HMM* offre plusieurs pistes de recherche ultérieure sur la modélisation et l'estimation de la volatilité des séries financières. Allant par exemple de l'intégration des sauts dans l'équation de la variable latente à la considération de modèles à volatilité stochastique avec changement de régimes. En effet, à notre point de vue, il serait plus facile d'étendre la méthodologie développée dans le cadre d'un modèle à volatilité stochastique avec changement de régimes que celle du *filtre particulière continu*. Une étude de la performance de ces deux généralisations dans ce cadre pourrait donc être menée.

Bibliographie

- Abramowitz, M. et Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 9. Dover, New York.
- Andersen, T. (1994). Stochastic autoregressive volatility : a framework for volatility modeling. *Mathematical finance*, 4(2) :75–102.
- Augustyniak, M. (2017). *Méthodes d'inférence pour les modèles à chaîne de Markov cachée : Notes de cours (STT6705V)*. Département de mathématiques et de statistique, Université de Montréal.
- Augustyniak, M., Bauwens, L., et Dufays, A. (2017). A new approach to volatility modeling : the factorial hidden Markov volatility model. *Journal of Business & Economic Statistics* (accepted).
- Augustyniak, M., Boudreault, M., et Morales, M. (2018). Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. *Methodology and Computing in Applied Probability*, 20(1) :165–188.
- Bartolucci, F. et De Luca, G. (2001). Maximum likelihood estimation of a latent variable time-series model. *Applied Stochastic Models in Business and Industry*, 17(1) :5–17.
- Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177–181.
- Black, F. et Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3) :637–654.
- Blight, B. (1974). Recursive solutions for the estimation of a stochastic parameter. *Journal of the American Statistical Association*, 69(346) :477–481.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3) :307–327.

- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3) : 542–547.
- Bollerslev, T. (2010). *Glossary to ARCH (GARCH)*. In Bollerslev, T., Russell, J., and Watson, M., editors, *Volatility and Time Series Econometrics : Essays in Honor of Robert Engle*. Oxford University Press, USA.
- Bollerslev, T., Chou, R. Y., et Kroner, K. F. (1992). ARCH modeling in finance : A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1-2) :5–59.
- Bollerslev, T., Engle, R., et Nelson, D. (1994). ARCH models. *Handbook of Econometrics*, 4 :2959–3038. Amsterdam : North-Holland.
- Broto, C. et Ruiz, E. (2004). Estimation methods for stochastic volatility models : a survey. *Journal of Economic Surveys*, 18(5) :613–649.
- Cappé, O., Moulines, E., et Rydén, T. (2005). Inference in Hidden Markov Models. *Springer Series in Statistics*. Springer.
- Carnero, M., Peña, D., et Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, 2(2) :319–342.
- Chib, S., Nardari, F., et Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2) :281–316.
- Christie, A. (1982). The stochastic behavior of common stock variances : Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4) :407–432.
- Clements, A., White, S., et Hurn, S. (2004). *Discretised Non-Linear Filtering for Dynamic Latent Variable Models : with Application to Stochastic Volatility*. School of Economics and Finance, Queensland University of Technology.
- Cont, R. (2001). Empirical properties of asset returns : stylized facts and statistical issues. *Quantitative Finance*, 1(2) :223–236.
- Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric reviews*, 31(3) :245–296.
- Crişan, D., Del Moral, P., et Lyons, T. J. (1999). Interacting particle systems approximations of the Kushner-Stratonovitch equation. *Advances in Applied Probability*, 31(3) :819–838.
- Crişan, D. et Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on signal processing*, 50(3) :736–746.

- Danielsson, J. (1994). Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics*, 64(1) :375–400.
- Danielsson, J. et Richard, J.-F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8(S1) :S153–S173.
- Del Moral, P. (1996). Non-linear filtering : interacting particle resolution. *Markov processes and related fields*, 2(4) :555–581.
- Del Moral, P. et Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Seminaire de probabilites XXXIV*, 1–145. Springer.
- Doucet, A., Godsill, S., et Andrieu, C. (2000). On sequential Monte Carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3) :197–208.
- Doucet, A. et Johansen, A. (2009). A tutorial on particle filtering and smoothing : Fifteen years later. In D. Crisan and B. Rozovsky (eds.), *The Oxford Handbook of Nonlinear Filtering, Oxford Handbooks in Mathematics*, 656–704 Oxford University Press, New York.
- Durbin, J. et Koopman, S. J. (2012). Time Series Analysis by State Space Methods. *Volume 38 of Oxford Statistical Science Series*, OUP Oxford.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4) :987–1007.
- Fama, E. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1) :34–105.
- Fridman, M. et Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics*, 16(3) :284–291.
- Ghysels, E. et Jasiak, J. (1994). Stochastic volatility and time deformation : an application of trading volume and leverage effects. *Paper presented at the Western Finance Association Meetings, Santa Fe*.
- Ghysels, E., Harvey, A. et Renault, E. (1996). Stochastic volatility. In GS Maddala and CR Rao (eds), *Chapter 14 Handbook of statistics*, 14 :119–191.
- Glosten, L., Jagannathan, R., et Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5) :1779–1801.

- Gordon, N., Salmond, D., et Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F, Radar and Signal Processing*, 140(2) :107–113.
- Grimmett, G. et Stirzaker, D. (2001). *Probability and random processes*, Oxford university press, New York, third edition.
- Hansen, L. (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica*, 50(4) : 1029-1054.
- Harvey, A., Ruiz, E., et Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2) :247–264.
- Harvey, A. et Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, 14(4) :429–434.
- Hull, J. et White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2) :281–300.
- Jacquier, E., Polson, N., et Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics*, 12 :371–417.
- Jacquier, E., Polson, N., et Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1) :185–212.
- Kalman, R. et Bucy, R. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME-Journal of Basic Engineering*, Series D, 83 :95–108.
- Kearns, P. et Pagan, A. (1997). Estimating the density tail index for financial time series. *The Review of Economics and Statistics*, 79(2) :171–175.
- Kim, S., Shephard, N., et Chib, S. (1998). Stochastic volatility : likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3) :361–393.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400) :1032–1041.
- Langrock, R., MacDonald, I., et Zucchini, W. (2012). Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models. *Journal of Empirical Finance*, 19(1) :147–161.
- Liesenfeld, R. et Jung, R. (2000). Stochastic volatility models : conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics*, 15(2) :137–160.

- Lystig, T. C. et Hughes, J. (2002). Exact computation of the observed information matrix for hidden Markov models. *Journal of Computational and Graphical Statistics*, 11(3) :678–689.
- Malik, S. et Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximization. *Journal of Econometrics*, 165(2) :190–209.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4) :394–419.
- Mandelbrot, B. (1971). When can price be arbitrated efficiently? a limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, MIT Press, 53(3) : 225-236.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1) :77–91.
- Melino, A. et Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2) :239–265.
- Nelson, D. (1988). *The time series behavior of stock market volatility and returns*. PhD thesis, Massachusetts Institute of Technology.
- Osborne, M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2) :145–173.
- Pitt, M., Malik, S., et Doucet, A. (2014). Simulated likelihood inference for stochastic volatility models using continuous particle filtering. *Annals of the Institute of Statistical Mathematics*, 66(3) :527–552.
- Polson, N., Stroud, J., et Müller, P. (2008). Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 70(2) :413–428.
- Ristic, B., Arulampalam, S., et Gordon, N. (2004). *Beyond the Kalman filter : Particle filters for tracking applications*. Artech House, London, 3 :1077–2626.
- Robert, C. et Casella, G. (2004). Monte Carlo Statistical Methods. *Springer Texts in Statistics*. Springer-Verlag, New York, second edition.
- Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from august 23, 1976 through august 31, 1978. *The Journal of Finance*, 40(2) :455–480.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63(1) :289–306.

- Sandmann, G. et Koopman, S. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87(2) :271–301.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models in Econometrics, Finance and Other Fields*, (edited by D.R. Cox, David V. Hinkley and Ole E. Barndorff-Neilsen), London : Chapman & Hall, 1-67.
- Shephard, N. (2000). Discussion of Durbin and Koopman (2000). (Time series analysis of non-Gaussian observations based on state-space models from both classical and Bayesian perspectives). *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 62(1) :3–56.
- Shephard, N. (2005). Stochastic volatility : selected readings. Oxford University Press.
- Shumway, R. H. et Stoffer, D. S. (2011). *Time Series Analysis and Its Applications : with R examples*. Springer Science & Business Media, New York, third edition.
- So, M., Li, W., et Lam, K. (2002). A threshold stochastic volatility model. *Journal of Forecasting*, 21(7) :473–500.
- Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961-75. *Time series analysis : theory and practice*, 1 :203–226.
- Taylor, S. (1986). Modelling financial time series. (John Wiley : Chichester).
- Taylor, S. (1994). Modeling stochastic volatility : A review and comparative study. *Mathematical finance*, 4(2) :183–204.
- Tsay, R. S. (2010). *Analysis of financial time series*, Wiley Series in Probability and Statistics. John Wiley & Sons., New Jersey, third edition.
- Wagner, N. et Marsh, T. (2005). Measuring tail thickness under GARCH and an application to extreme exchange rate changes. *Journal of Empirical Finance*, 12(1) :165–185.
- Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2) :165–178.
- Zucchini, W. et MacDonald, I. (2009). *Hidden Markov Models for Time Series : An Introduction Using R*, volume 150. CRC Press.

Annexe A

PREUVE DE LA PROPOSITION 2.2.1

Soit le modèle AR(1) gaussien bruité présenté à la section (2.2.1) :

$$p(y_t | h_t) \sim N(h_t, \sigma_y^2),$$

$$p(h_t | h_{t-1}) \sim N(\phi h_{t-1}, \sigma_\eta^2),$$

$$p(h_0) \sim N\left(0, \frac{\sigma_\eta^2}{1 - \phi^2}\right).$$

Cette annexe offre une preuve de la proposition (2.2.1) par évaluation directe des intégrales des étapes de prédictions et de mise à jour résolvant le problème de filtrage (2.1.3).

On adopte ici la notation introduite dans [Blight \(1974\)](#) :

$$\{x; \mu, \sigma^2\} := \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} = \exp\left\{-\frac{1}{2\sigma^2}(\mu - x)^2\right\} =: \{\mu; x, \sigma^2\} \quad (\text{A.1})$$

Posons $A = -\frac{1}{2\sigma_1^2}(x - \mu_1)^2 - \frac{1}{2\sigma_2^2}(x - \mu_2)^2$, on a :

$$\begin{aligned} A &= -\frac{1}{2} \left\{ x^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - 2x \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) + \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} \right) \right\}, \\ &= -\frac{1}{2} \left\{ \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \left(x - \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 - \left(\frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) + \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} \right) \right\}, \\ &= -\frac{1}{2} \left\{ \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left(x - \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 + \frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} - \left(\frac{\mu_1^2}{\sigma_1^4} + \frac{\mu_2^2}{\sigma_2^4} + \frac{2\mu_1\mu_2}{\sigma_1^2\sigma_2^2} \right) \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right\}, \\ &= -\frac{1}{2} \left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left(x - \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 \right) \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{2} \left(\frac{1}{\sigma_1^2 + \sigma_2^2} \left(\mu_1^2 + \frac{\sigma_2^2}{\sigma_1^2} \mu_1^2 + \mu_2^2 + \frac{\sigma_1^2}{\sigma_2^2} \mu_2^2 - \frac{\sigma_2^2}{\sigma_1^2} \mu_1^2 - \frac{\sigma_1^2}{\sigma_2^2} \mu_2^2 - 2\mu_1\mu_2 \right) \right), \\
& = -\frac{1}{2} \left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left(x - \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 \right) - \frac{1}{2} \left(\frac{1}{\sigma_1^2 + \sigma_2^2} (\mu_1 - \mu_2)^2 \right),
\end{aligned}$$

et donc,

$$\{x; \mu_1, \sigma_1^2\} \times \{x; \mu_2, \sigma_2^2\} = \left\{ x; \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right\} \times \{\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2\} \quad (\text{A.2})$$

Maintenant, en utilisant les résultats (A.1) et (A.2), nous avons les itérations suivantes.

En $t = 0$ nous savons que la distribution initiale de la variable d'état H_0 est donnée par :
 $p(h_0) \sim N(\hat{h}_{0|0}, P_{0|0})$, avec

$$\hat{h}_{0|0} = 0, \quad \text{et} \quad P_{0|0} = \frac{\sigma_\eta^2}{1 - \phi^2}.$$

En partant donc de $p(h_0)$, le problème de filtrage pour ce modèle se résout en itérant les étapes suivantes pour $t = 1, \dots, T$.

- $p(h_t | y_{1:t-1})$

$$\begin{aligned}
p(h_t | y_{1:t-1}) &= \int_{\mathbb{R}} p(h_t | h_{t-1}) p(h_{t-1} | y_{1:t-1}) dh_{t-1}, \\
&= \frac{1}{\sigma_\eta \phi P_{t-1|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \{h_t; \phi h_{t-1}, \sigma_\eta^2\} \times \{h_{t-1}; \hat{h}_{t-1|t-1}, P_{t-1|t-1}\} dh_{t-1}, \\
&= \frac{1}{\sigma_\eta \phi P_{t-1|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \{\phi h_{t-1}; h_t, \sigma_\eta^2\} \times \{\phi h_{t-1}; \hat{\phi} \hat{h}_{t-1|t-1}, \phi^2 P_{t-1|t-1}\} dh_{t-1}, \\
&= \frac{1}{\sigma_\eta P_{t-1|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \left\{ \phi h_{t-1}; \frac{h_t}{\sigma_\eta^2} + \frac{\phi \hat{h}_{t-1|t-1}}{\phi^2 P_{t-1|t-1}}, \frac{\phi^2 \sigma_\eta^2 P_{t-1|t-1}}{\phi^2 P_{t-1|t-1} + \sigma_\eta^2} \right\} \\
&\quad \times \{h_t; \hat{\phi} \hat{h}_{t-1|t-1}, \phi^2 P_{t-1|t-1} + \sigma_\eta^2\} dh_{t-1}, \\
&= \frac{\{h_t; \hat{\phi} \hat{h}_{t-1|t-1}, \phi^2 P_{t-1|t-1} + \sigma_\eta^2\}}{\sqrt{2\pi} (\phi^2 P_{t-1|t-1} + \sigma_\eta^2)} \times \frac{\left(\frac{\phi^2 \sigma_\eta^2 P_{t-1|t-1}}{\phi^2 P_{t-1|t-1} + \sigma_\eta^2} \right)^{-\frac{1}{2}}}{\sqrt{2\pi}}
\end{aligned}$$

$$\begin{aligned}
& \times \int_{\mathbb{R}} \left\{ \phi h_{t-1}; \frac{h_t}{\sigma_\eta^2} + \frac{\phi \hat{h}_{t-1|t-1}}{\phi^2 P_{t-1|t-1}}, \frac{\phi^2 \sigma_\eta^2 P_{t-1|t-1}}{\phi^2 P_{t-1|t-1} + \sigma_\eta^2} \right\} dh_{t-1}, \\
& = \frac{\{h_t; \phi \hat{h}_{t-1|t-1}, \phi^2 P_{t-1|t-1} + \sigma_\eta^2\}}{\sqrt{2\pi(\phi^2 P_{t-1|t-1} + \sigma_\eta^2)}}, \\
& \sim N(\hat{h}_{t|t-1}, P_{t|t-1}),
\end{aligned}$$

avec

$$\hat{h}_{t|t-1} = \mathbb{E}[h_t | y_{1:t-1}] = \phi \hat{h}_{t-1|t-1}, \quad (\text{A.3})$$

$$P_{t|t-1} = \text{Var}[h_t | y_{1:t-1}] = \phi^2 P_{t-1|t-1} + \sigma_\eta^2. \quad (\text{A.4})$$

• $p(h_t | y_{1:t})$

$$\begin{aligned}
p(h_t | y_{1:t}) &= \frac{p(y_t | h_t) p(h_t | y_{1:t-1})}{p(y_t | y_{1:t-1})}, \\
&\propto p(y_t | h_t) p(h_t | y_{1:t-1}), \\
&= \frac{1}{\sigma_\epsilon P_{t|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \{h_t; y_t, \sigma_\epsilon^2\} \times \{h_t; \hat{h}_{t|t-1}, P_{t|t-1}\}, \\
&= \frac{\{y_t; \hat{h}_{t|t-1}, P_{t|t-1} + \sigma_\epsilon^2\}}{\sqrt{2\pi(P_{t|t-1} + \sigma_\epsilon^2)}} \times \frac{\left(\frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}\right)^{-\frac{1}{2}}}{\sqrt{2\pi}} \left\{ h_t; \frac{y_t}{\sigma_\epsilon^2} + \frac{\hat{h}_{t|t-1}}{P_{t|t-1}}, \frac{P_{t|t-1} + \sigma_\epsilon^2}{\sigma_\epsilon^2 + P_{t|t-1}} \right\}, \\
&= \frac{\left(\frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}\right)^{-\frac{1}{2}}}{\sqrt{2\pi}} \left\{ h_t; \frac{\hat{h}_{t|t-1} \sigma_\epsilon^2 + y_t P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}, \frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2} \right\}, \\
&\sim N(\hat{h}_{t|t}, P_{t|t}),
\end{aligned}$$

avec

$$\begin{aligned}
\hat{h}_{t|t} = \mathbb{E}[h_t | y_{1:t}] &= \frac{\hat{h}_{t|t-1} \sigma_\epsilon^2 + y_t P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}, \\
&= \frac{\hat{h}_{t|t-1} (P_{t|t-1} + \sigma_\epsilon^2) - \hat{h}_{t|t-1} P_{t|t-1} + y_t P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}, \\
&= \hat{h}_{t|t-1} + P_{t|t-1} (P_{t|t-1} + \sigma_\epsilon^2)^{-1} (y_t - \hat{h}_{t|t-1}),
\end{aligned}$$

$$= \hat{h}_{t|t-1} + P_{t|t-1}\Sigma_t^{-1}(y_t - \hat{y}_t), \quad \text{\`a partir de (A.7) et (A.8),} \quad (\text{A.5})$$

$$\begin{aligned} P_{t|t} = \text{Var}[h_t | y_{1:t}] &= \frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}, \\ &= P_{t|t-1} \frac{\sigma_\epsilon^2}{P_{t|t-1} + \sigma_\epsilon^2}, \\ &= P_{t|t-1} \left(1 - \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2} \right), \\ &= P_{t|t-1} - P_{t|t-1}\Sigma_t^{-1}P_{t|t-1}, \quad \text{\`a partir de (A.8).} \end{aligned} \quad (\text{A.6})$$

- $p(y_t | y_{1:t-1})$

En $t = 1$, nous avons pour $p(y_1 | y_{1:0})$ la distribution suivante :

$$\begin{aligned} p(y_1) &= \int_{\mathbb{R}} p(y_1 | h_1) p(h_1) dh_1, \\ &= \frac{1}{\sigma_\epsilon P_{1|0}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \{h_1; y_1, \sigma_\epsilon^2\} \times \{h_1; \hat{h}_{1|0}, P_{1|0}\} dh_1, \\ &= \frac{1}{\sigma_\epsilon P_{1|0}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \left\{ h_1; \frac{y_1}{\sigma_\epsilon^2} + \frac{\hat{h}_{1|0}}{P_{1|0}}, \frac{\sigma_\epsilon^2 P_{1|0}}{P_{1|0} + \sigma_\epsilon^2} \right\} \times \{y_1; \hat{h}_{1|0}, P_{1|0} + \sigma_\epsilon^2\} dh_1, \\ &= \frac{\{y_1; \hat{h}_{1|0}, P_{1|0} + \sigma_\epsilon^2\}}{\sqrt{2\pi}(P_{1|0} + \sigma_\epsilon^2)} \times \frac{\left(\frac{\sigma_\epsilon^2 P_{1|0}}{P_{1|0} + \sigma_\epsilon^2} \right)^{-\frac{1}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} \left\{ h_1; \frac{y_1}{\sigma_\epsilon^2} + \frac{\hat{h}_{1|0}}{P_{1|0}}, \frac{\sigma_\epsilon^2 P_{1|0}}{P_{1|0} + \sigma_\epsilon^2} \right\} dh_1, \\ &= \frac{\{y_1; \hat{h}_{1|0}, P_{1|0} + \sigma_\epsilon^2\}}{\sqrt{2\pi}(P_{1|0} + \sigma_\epsilon^2)}, \\ &\sim N(\hat{y}_1, \Sigma_1), \end{aligned}$$

avec $\hat{y}_1 = \hat{h}_{1|0}$ et $\Sigma_1 = P_{1|0} + \sigma_\epsilon^2$. Maintenant pour $t = 2, \dots$, on a :

$$\begin{aligned} p(y_t | y_{1:t-1}) &= \int_{\mathbb{R}} p(y_t | h_t) p(h_t | y_{1:t-1}) dh_t, \\ &= \frac{1}{\sigma_\epsilon P_{t|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \{h_t; y_t, \sigma_\epsilon^2\} \times \{h_t; \hat{h}_{t|t-1}, P_{t|t-1}\} dh_t, \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma_\epsilon P_{t|t-1}^{\frac{1}{2}} (\sqrt{2\pi})^2} \int_{\mathbb{R}} \left\{ h_t; \frac{y_t}{\sigma_\epsilon^2} + \frac{\hat{h}_{t|t-1}}{P_{t|t-1}}, \frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2} \right\} \times \{y_t; \hat{h}_{t|t-1}, P_{t|t-1} + \sigma_\epsilon^2\} dh_t, \\
&= \frac{\{y_t; \hat{h}_{t|t-1}, P_{t|t-1} + \sigma_\epsilon^2\}}{\sqrt{2\pi}(P_{t|t-1} + \sigma_\epsilon^2)} \times \frac{\left(\frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}\right)^{-\frac{1}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} \left\{ h_t; \frac{y_t}{\sigma_\epsilon^2} + \frac{\hat{h}_{t|t-1}}{P_{t|t-1}}, \frac{\sigma_\epsilon^2 P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2} \right\} dh_t, \\
&= \frac{\{y_t; \hat{h}_{t|t-1}, P_{t|t-1} + \sigma_\epsilon^2\}}{\sqrt{2\pi}(P_{t|t-1} + \sigma_\epsilon^2)}, \\
&\sim N(\hat{y}_t, \Sigma_t),
\end{aligned}$$

avec

$$\hat{y}_t = \mathbb{E}[Y_t | y_{1:t-1}] = \hat{h}_{t|t-1}, \quad (\text{A.7})$$

$$\Sigma_t = \text{Var}[Y_t | y_{1:t-1}] = P_{t|t-1} + \sigma_\epsilon^2. \quad (\text{A.8})$$

Annexe B

INTRODUCTION AUX MODÈLES À CHAÎNE DE MARKOV CACHÉE

Cette annexe vise à faire une brève introduction au modèle à chaîne de Markov cachée qui est un élément central de ce mémoire.

B.1. CHAÎNE DE MARKOV À ESPACE D'ÉTATS FINI

La notion de chaîne de Markov à espace d'états fini est un élément important dans la conception même des modèles à chaîne de Markov cachée. Nous rappelons ici quelques propriétés des chaînes de Markov à espace d'états fini. Le chapitre 6 du livre [Grimmett et Stirzaker \(2001\)](#) peut être consulté pour une présentation plus détaillée sur le sujet.

Définition B.1.1. *Une suite $\{X_t\}$ de variables aléatoires à valeurs dans un espace d'états fini $\mathcal{S} \in \mathbb{R}^N$ est une **chaîne de Markov à temps discret** si elle possède la **propriété de Markov** :*

$$\forall t : \mathbb{P}[X_{t+1} = i_{t+1} \mid X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_1 = i_1] = \mathbb{P}[X_{t+1} = i_{t+1} \mid X_t = i_t]$$

pour toute suite $i_{t+1}, i_t, \dots, i_1 \in \mathcal{S}$.

Une chaîne de Markov est donc un processus sans mémoire, liant le futur au passé uniquement au travers de l'information présente. La notion de **probabilités de transition** sert à formaliser et quantifier ce lien.

Définition B.1.2. *La **matrice de transition** d'une chaîne de Markov, à la date t notée Γ_t , est la matrice contenant, au croisement de la ligne i et de la colonne j , la probabilité de passer de l'état i à l'état j : $\gamma_{ij}(t)$.*

Ainsi :

$$\gamma_{ij}(t) := \mathbb{P}(X_{t+1} = j \mid X_t = i),$$

avec $0 \leq \gamma_{ij}(t) \leq 1$, $i, j \in \{i_1, \dots, i_N\} = \mathcal{S}$ et $t = 1, \dots, T - 1$.

Lorsque les probabilités de transitions $\gamma_{ij}(t)$ ne dépendent pas de t , la chaîne de Markov est dite homogène, et la matrice de transition Γ s'exprime donc pour tout t :

$$\mathbf{\Gamma} := \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \gamma_{ij} & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{pmatrix}.$$

Définition B.1.3. *Lorsque chaque état de la chaîne de Markov est accessible en un ou plusieurs pas à partir de n'importe quels autres états, cette chaîne de Markov ainsi que sa matrice de transition sont dites **irréductible**.*

L'approximation de la vraisemblance des modèles à volatilité stochastique par le *filtre HMM* a donné lieu à une matrice de transition irréductible et homogène dans le cas du modèle SV_0 et des matrices de transitions irréductibles et non-homogènes pour les modèles SVL et SVLJ.

On définit de manière analogue la matrice de transition en k pas de temps $\mathbf{\Gamma}^{(k)}$ comme la matrice contenant les probailités :

$$\gamma_{ij}^{(k)}(t) := \mathbb{P}(X_{t+k} = j \mid X_t = i), \quad i, j = i_1, \dots, i_N.$$

Cette matrice satisfait l'équation de Chapman-Kolmogorov :

$$\mathbf{\Gamma}^{(k+l)} = \mathbf{\Gamma}^{(k)}\mathbf{\Gamma}^{(l)}, \quad k \in \mathbb{N} \text{ et } l \in \mathbb{N}.$$

et donc,

$$\mathbf{\Gamma}^{(k)} = \mathbf{\Gamma}^k, \quad k \in \mathbb{N}. \tag{B.1}$$

En plus de la matrice de transition, il suffit de donner la **distribution initiale** pour que la chaîne de Markov soit complètement caractérisée. La distribution initiale d'une chaîne de Markov est donnée par le vecteur représentant sa distribution marginale en $t = 1$ dans chacun des N états :

$$\mathbf{u}(0) := (\mathbb{P}(X_1 = 1), \dots, \mathbb{P}(X_1 = N)).$$

La distribution marginale à une date $k \geq 1$ est alors donnée par la relation :

$$\mathbf{u}(k) = \mathbf{u}(1 + (k - 1)) = \mathbf{u}(1)\mathbf{\Gamma}^{k-1}, \quad k = 1, 2, \dots \quad (\text{B.2})$$

Définition B.1.4. *Le vecteur $\boldsymbol{\pi} := (\pi_1, \dots, \pi_N)$ est une distribution stationnaire de la chaîne de Markov de matrice de transition $\mathbf{\Gamma}$ si les trois conditions suivantes sont satisfaites :*

$$\begin{aligned} (1) &= \pi_j \geq 0, \quad j = 1, \dots, N, \\ (2) &= \boldsymbol{\pi} \mathbf{1}_N = 1, \quad (\text{ou } \sum_{j=1}^N \pi_j = 1) \\ (3) &= \boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{\Gamma}. \quad (\text{ou } \pi_j = \sum_{i=1}^N \pi_i \gamma_{ij}) \end{aligned}$$

Il est de pratique courante de prendre comme distribution initiale d'une chaîne de Markov sa distribution stationnaire lorsqu'elle existe¹. Dans le cas échéant,

$$\mathbf{u}(k) = \mathbf{u}(1)\mathbf{\Gamma}^{k-1} = \boldsymbol{\pi}\mathbf{\Gamma}^{k-1}, \quad k = 1, 2, \dots \quad (\text{B.3})$$

Une chaîne de Markov ayant sa distribution stationnaire comme distribution initiale est un processus stationnaire au sens strict.

B.2. MODÈLES À CHAÎNE DE MARKOV CACHÉE

Définition B.2.1. *Un modèle à chaîne de Markov cachée (HMM : hidden Markov model) est un processus stochastique à temps discret bivarié $\{X_t\} = \{(Y_t, H_t)\}$ partiellement observé. Dans cette représentation, le processus $\{Y_t\}$ est observé et sa distribution est gouvernée par le processus non-observé ou latent $\{H_t\}$, qui est une chaîne de Markov à valeurs dans un espace fini $\mathcal{S} \in \mathbb{R}^N$, d'où l'appellation chaîne de Markov cachée. Dans sa forme la plus simple, on suppose que les observations sont générées à travers un canal sans mémoire, c'est-à-dire que conditionnellement aux valeurs prises par $\{H_t\}$, $\{Y_t\}$ est une suite de variables aléatoires indépendantes, et que la distribution de chaque observation y_t ne dépend que de l'état h_t du processus $\{H_t\}$ à cette instant.*

1. L'existence et l'unicité de la distribution stationnaire d'une chaîne de Markov irréductible est une conséquence du théorème de Perron-Frobenius.

Formellement, le modèle se spécifie comme suit :

$$(HMM) : \begin{cases} H_1 \sim p(h_1), \\ H_t | (Y_{1:t-1} = y_{1:t-1}, H_{1:t-1} = h_{1:t-1}) \sim p(h_t | h_{t-1}), \quad t = 1, 2, \dots, \\ Y_t | (Y_{1:t-1} = y_{1:t-1}, H_{1:t} = h_{1:t}) \sim p(y_t | h_t), \quad t \in \mathbb{N}. \end{cases} \quad (\text{B.1})$$

On peut facilement remarquer que le processus bivarié $\{X_t\} = \{(Y_t, H_t)\}$ est aussi une chaîne de Markov,

$$\begin{aligned} p(x_t | x_{1:t-1}) &= p(y_t, h_t | y_{1:t-1}, h_{1:t-1}) = p(h_t | y_{1:t-1}, h_{1:t-1}) p(y_t | y_{1:t-1}, h_{1:t-1}, h_t) \\ &= p(h_t | y_{t-1}, h_{t-1}) p(y_t | y_{t-1}, h_{t-1}, h_t) \\ &= p(y_t, h_t | y_{t-1}, h_{t-1}) = p(x_t | x_{t-1}) \end{aligned}$$

Toutefois, comme le montre le contre-exemple à la page 39 de [Zucchini et MacDonald \(2009\)](#) la distribution du processus observé $\{Y_t\}$ dépend généralement de toutes l'historique $\{h_{t-1}, \dots, h_1\}$ du processus latent $\{H_t\}$.

La distribution marginale du processus observé $\{Y_t\}$ peut être obtenue à partir de (B.1) et une simple application de la loi des probabilités totales,

$$\begin{aligned} p(y_t) &= \sum_{i=1}^N \mathbb{P}(H_t = i) p(y_t | H_t = i) \\ &= \sum_{i=1}^N u_i(t) p_i(y_t) \\ &= (\mathbf{u}_1(t), \dots, \mathbf{u}_N(t)) \begin{pmatrix} p_1(y_t) & & 0 \\ & \ddots & \\ 0 & & p_N(y_t) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \mathbf{u}(t) \mathbf{P}(y_t) \mathbf{1}_N, \end{aligned} \quad (\text{B.2})$$

où $\mathbf{u}(t) := (u_1(t), \dots, u_N(t))$ avec $u_i(t) := \mathbb{P}(H_t = i)$ pour $i = 1, \dots, N$. $\mathbf{P}(y_t)$ la matrice diagonale avec sur la diagonale les probabilités $p_i(y_t)$ avec $p_i(y_t) := p(y_t | h_t = i)$ pour $i = 1, \dots, N$ et $\mathbf{1}'_N = (1, \dots, 1)$.

Les deux propositions suivantes nous permettront d'exprimer la vraisemblance du processus observé $\{Y_t\}$. Soit $1 \leq t_1 \leq t_2 \leq \dots \leq t_n$, et notons

$$\mathbf{u}_{t_1, \dots, t_n}(y_{t_1}, \dots, y_{t_n}) := \left(p(y_{t_1}, \dots, y_{t_n}, H_{t_n} = 1), \dots, p(y_{t_1}, \dots, y_{t_n}, H_{t_n} = N) \right)$$

Proposition B.2.1.

$$\mathbf{u}_{t_1, \dots, t_n}(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}_{t_1, \dots, t_{n-1}}(y_{t_1}, \dots, y_{t_{n-1}}) \mathbf{\Gamma}^{t_n - t_{n-1}} \mathbf{P}(y_{t_n})$$

DÉMONSTRATION. Pour $H_{t_n} = j$, $j = 1, \dots, N$ nous avons successivement :

$$\begin{aligned} p(y_{t_1}, \dots, y_{t_n}, H_{t_n} = j) &= \sum_{i=1}^N p(y_{t_1}, \dots, y_{t_{n-1}}, y_{t_n}, H_{t_{n-1}} = i, H_{t_n} = j) \\ &= \sum_{i=1}^N p(y_{t_1}, \dots, y_{t_{n-1}}, H_{t_{n-1}} = i) \mathbb{P}(H_{t_n} = j | H_{t_{n-1}} = i) p(y_{t_n} | H_{t_n} = j) \\ &= \sum_{i=1}^N \left(p(y_{t_1}, \dots, y_{t_{n-1}}, H_{t_{n-1}} = i) \right) \left(\gamma_{ij}^{(t_n - t_{n-1})} p_j(y_{t_n}) \right), \end{aligned} \quad (\text{B.3})$$

on en déduit donc

$$\mathbf{u}_{t_1, \dots, t_n}(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}_{t_1, \dots, t_{n-1}}(y_{t_1}, \dots, y_{t_{n-1}}) \mathbf{\Gamma}^{t_n - t_{n-1}} \mathbf{P}(y_{t_n})$$

□

Proposition B.2.2.

$$p(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}(t_1) \mathbf{P}(y_{t_1}) \mathbf{\Gamma}^{t_2 - t_1} \mathbf{P}(y_{t_2}) \dots \mathbf{\Gamma}^{t_n - t_{n-1}} \mathbf{P}(y_{t_n}) \mathbf{1}_N$$

DÉMONSTRATION. Notons d'abord que

$$p(y_{t_1}, \dots, y_{t_n}) = \sum_{i=1}^N p(y_{t_1}, \dots, y_{t_n}, H_{t_n} = i) = \mathbf{u}_{t_1, \dots, t_n}(y_{t_1}, \dots, y_{t_n}) \mathbf{1}_N.$$

Maintenant, comme

$$p(y_{t_1}, H_{t_1} = i) = \mathbb{P}(H_{t_1} = i) p(y_{t_1} | H_{t_1} = i) = u_i(t_1) p_i(y_{t_1}),$$

nous avons

$$\mathbf{u}_{t_1}(y_{t_1}) = \left(p(y_{t_1}, H_{t_1} = 1), \dots, p(y_{t_1}, H_{t_1} = N) \right) = \mathbf{u}(t_1) \mathbf{P}(y_{t_1}).$$

Il vient alors de la proposition (B.2.1),

$$\mathbf{u}_{t_1, \dots, t_n}(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}(t_1) \mathbf{P}(y_{t_1}) \mathbf{\Gamma}^{t_2 - t_1} \mathbf{P}(y_{t_2}) \dots \mathbf{\Gamma}^{t_n - t_{n-1}} \mathbf{P}(y_{t_n})$$

et le résultat s'ensuit donc

$$p(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}(t_1) \mathbf{P}(y_{t_1}) \mathbf{\Gamma}^{t_2 - t_1} \mathbf{P}(y_{t_2}) \dots \mathbf{\Gamma}^{t_n - t_{n-1}} \mathbf{P}(y_{t_n}) \mathbf{1}_N.$$

□

Vraisemblance du modèle

Dans le cas particulier où $(t_1, t_2, \dots, t_n) = (1, 2, \dots, T)$, une conséquence immédiate de la proposition (B.2.2) est que la vraisemblance d'un échantillon $y_{1:T}$ provenant de ce modèle à chaîne de Markov cachée serait donnée par :

$$p(\mathbf{y}_{1:T}) = p(y_{t_1}, \dots, y_{t_n}) = \mathbf{u}(1) \mathbf{P}(y_1) \mathbf{\Gamma} \mathbf{P}(y_2) \dots \mathbf{\Gamma} \mathbf{P}(y_T) \mathbf{1}_N, \quad (\text{B.4})$$

et peut être calculer récursivement à l'aide de l'**algorithme de filtrage avant** (*forward filtering algorithm*) comme suit :

$$\text{Étape 1. : } \boldsymbol{\alpha}_1 = \mathbf{u}(1) \mathbf{P}(y_1)$$

$$\text{Étape 2. : } \boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \mathbf{P}(y_t), \quad t = 2, \dots, T$$

$$\text{Étape 3. : } p(y_{1:T}) = \boldsymbol{\alpha}_T \mathbf{1}_N.$$

Le filtre d'Hamilton peut être généralisé pour des chaîne de Markov cachée plus complexe, voir la section 4.6.5 des notes de cours ([Augustyniak, 2017](#), *Méthode d'inférence pour les modèles à chaîne de Markov cachée (STT6705V)*) et [Augustyniak et al. \(2018\)](#) pour des applications.

Plusieurs autres résultats utiles pour l'inférence, la prévision et le filtrage des états du modèle peuvent aussi être dérivés relativement sans grandes difficultés. Par exemple, l'**algorithme de Viterbi** peut être utilisé pour déterminer la séquence optimale du processus latent $\{h_t\}$ au vue des observations $y_{1:t}$. Le gradient et la matrice hessienne de la fonction de log-vraisemblance du modèle à chaîne de Markov cachée peuvent aussi être dérivés récursivement et de manière exacte lors du calcul de la log-vraisemblance en adaptant, à cet effet, l'algorithme d'Hamilton (2.2.2) (voir : [Lystig et Hughes \(2002\)](#) pour plus de détails).

Nous référons le lecteur aux notes de cours *Méthode d'inférence pour les modèles à chaîne de Markov cachée (STT6705V)* [Augustyniak \(2017\)](#) et au livre [Zucchini et MacDonald \(2009\)](#) pour une introduction sur le sujet et pour aller plus en profondeur voir le livre de [Cappé et al. \(2005\)](#).