Volterra processes and applications in finance

Elizabeth Zuniga

Thèse de doctorat

NNT : 2021UPASM015

Processus de Volterra et applications en finance

Volterra processes and applications in finance

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 574, mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées
Unité de recherche : Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry, 91037, Evry-Courcouronnes, France.
Référent : Université d'Évry-Val d'Essonne

Thèse présentée et soutenue à Paris-Saclay, le 17/06/2021, par

Elizabeth ZUÑIGA

Composition du Jury

Christa CUCHIERO Professeure des universités, Université de Vienne	Présidente
Aurélien ALFONSI Chercheur, CERMICS, École des Ponts, ParisTech	Rapporteur & Examinateur
Christian BAYER Chercheur, Institut Weierstrass de Berlin	Rapporteur & Examinateur
Shiqi SONG Professeur des universités, Université d'Évry, Université Paris-Saclay	Examinateur

Direction de la thèse

Etienne CHEVALIER Maître de conférences, Université d'Évry, Université Paris-Saclay	Directeur de thèse
Sergio PULIDO Maître de conférences, ENSIIE, Université Paris-Saclay	Co-Encadrant de thèse

université evry val-d'essonne

université PARIS-SACLAY

LaMME Laboratoire de Mathématiques et Modélisation d'Évry

Fondation mathématique FMJH Jacques Hadamard

“The beauty of mathematics only shows itself to more patient followers. ”

Maryam Mirzakhani

Acknowledgments

I would like to thank the following people, without whom I would not have been able to complete this research.

First, I would like to express my greatest and most sincere gratitude to my supervisors Etienne Chevalier and Sergio Pulido, I thank them deeply for their guidance, patience and their immense knowledge. You provided me with the tools that I needed to successfully complete this work.

I would also like to thank my reviewers Aurélien Alfonsi and Christian Bayer, who did me the great honor of reporting this research work. Their insightful comments strengthen the quality of this research. Likewise, all my gratitude to Christa Cuchiero and Shiqi Song for agreeing to be part of my dissertation committee.

I would like to offer my special thanks to all the members of the Laboratoire de Mathématiques et Modélisation d'Évry for a cherished time spent together in the laboratory. In particular, a very special thanks to Agathe Guilloux and Arnaud Gloter, directors of the laboratory, and to all my teachers for their devoted instruction.

I owe a deep sense of gratitude to AMARUN, this adventure of coming all the way to France would have not been possible without their help and all the colleagues, brilliant mathematicians and specially very good friends that nurtured and accompanied me in this journey.

My biggest thanks go to my family that has been with me in every step of the way. Your support and encouragement were invaluable to me and I would like to thank you for always being there for me and listening, no matter the time of the day. To my parents Rita and Juan, to my sisters Francisca, Janneth and Nathaly and to my partner Galo, you are my strength and your love is the greatest treasure I have in my life.

Contents

Introduction (Version française)

Dans l'ensemble des modèles mathématiques qui cherchent à reproduire la dynamique des marchés financiers, ceux qui permettent de valoriser et surtout couvrir des produits dits dérivés sont ceux qui ont sucité le plus d'intérêt. Le plus connu est le modèle de Black et Scholes (voir Black and Scholes [1973]). Dans ce modèle, le processus de prix d'un actif S est solution de l'équation différentielle stochastique suivante:

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

où μ est le paramètre de dérive (drift), σ celui de volatilité et B est un mouvement Brownien standard. Ce modèle a été et est très utilisé car les prix et stratégies de couverture d'options européennes y sont donnés par des formules explicites, facilement exploitables. Cependant, il ne correspond pas à certains aspects de la dynamique réelle des marchés financiers. En particulier, il a été observé que l'hypothèse de volatilité constante n'est pas valide. Pour surmonter ces limites du modèle de Black and Scholes, une nouvelle classe de modèles a été développée: les modèles à volatilité stochastique. On y suppose que le carré de la volatilité est lui-même un processus stochastique V qui, dans les premiers modèles, est aussi solution d'une équation stochastique différentielle

$$dV_t = \alpha(t, V)dt + \beta(t, V)dW_t,$$

où α et β sont des fonctions du temps et du processus V. W est un mouvement brownien corrélé avec B. Les premiers de ces modèles présentent en général des structures markoviennes et les processus en jeux sont des semi-martingales continues. Ces modèles à volatilité stochastique, dont le plus connu est celui de Heston (voir Heston [1993]), reproduisent correctement la structure par terme du skew à la monnaie pour de longues maturité mais ne parviennent pas à expliquer certains faits stylisés comme la convexité des surfaces de volatilité implicite, observée pour des maturités très courtes.

Plus précisément, notons $\sigma_{BS}(k,\tau)$ [1] la volatilité implicite d'une option dont le logarithme du quotient entre prix d'exercice et prix spot est k et le temps restant avant la maturité est τ. La structure par termes du skew à la monnaie est donnée par:

$$\xi(\tau) = \left|\frac{\partial \sigma_{BS}(k,\tau)}{\partial k}\right|_{k=0}$$

Des études empiriques ont montré une décroissance rapide du skew à la monnaie avec explosion pour τ tendant vers 0. Dans Fukasawa [2011], il est montré que, au moins pour de petites valeurs de τ la structure par terme issue de modèles à volatilité stochastique dont la dynamique est dirigée par un Brownien fractionnaire d'indice de Hurst H, est donné par $\xi(\tau) = \tau^{H-1/2}$. L'utilisation du Brownien fractionnaire a été initiée par les auteurs de Comte and Renault [1998] qui ont proposé de modéliser le logarithme de la variance en remplaçant la dynamique brownienne classique par celle d'un Brownien fractionnaire W^H, défini par:

$$W_t^H = \frac{1}{\Gamma(H+1/2)}\left(\int_{-\infty}^{0}\left((t-s)^{H-1/2} - (-s)^{H-1/2}\right)dW_s + \int_0^t (t-s)^{H-1/2} dW_s\right), \tag{1}$$

où l'indice de Hurst $H \in (0,1)$ gouverne la régularité des trajectoires de W^H. Les trajectoires de W^H sont en effet localement ölder iennes d'ordre strictement inférieur à H. In Comte and Renault [1998], l'indice de Hurst choisi est $H > 1/2$ pour rendre compte d'effet de mémoire longue dans la dynamique. Cependant, dans Gatheral et al. [2018] il est montré que le logarithme de la volatilité se comporte essentiellement comme un Brownien fractionnaire d'indice de Hurst de l'ordre de 0.1.

Ces observations ont contribuées à un intérêt croissant pour les modèles dits à volatilité rugueuse (voir, par exemple, Alòs et al. [2007], Fukasawa [2011], Bayer et al. [2016], Garnier and Solna [2017], Forde and Zhang [2017], Fukasawa [2017], Guennoun et al. [2018], Bayer et al. [2019], Euch et al. [2019], Fukasawa [2021]). Les modèles à volatilité rugueuse sont donc des modèles à volatilité stochastique tels que les trajectoires des processus de volatilité sont continues mais moins régulières que celles d'un mouvement brownien. Plus spécifiquement, on parlera de trajectoires rugueuse pour une régularité höldérienne d'ordre strictement inférieur à 1/2.

[1] $\sigma_{BS}(k,\tau)$ est le paramètre de volatilité du modèle de Black et Scholes tel que le prix d'une option de caractéristiques (k,τ) (prix d'exercice, maturité) soit dans le, modèle égal au prix de marché observé de l'option avec les mêmes caractéristiques.

De plus, les modèles rugueux semblent remarquablement adaptés à la reproduction des séries temporelles issues des données financières. Des nombreuses preuves statistiques, à différentes échelles de temps et dans beaucoup de marchés différents, montrent que les séries temporelles réalisées de volatilité varient plus rapidement que les trajectoires d'un mouvement Brownien (voir Bennedsen et al. [2021], Gatheral et al. [2018], Livieri et al. [2018], Fukasawa et al. [2019]).

Ces arguments statistiques sont aussi renforcés par des considérations micro-structurelles car les modèles à volatilité rugueuse apparaissent naturellement comme modèles limites de modèles micro-structurels dirigés par des processus auto-excitant comme les processus de Hawkes (voir Jaisson and Rosenbaum [2016], El Euch et al. [2018], Dandapani et al. [2019], Tomas and Rosenbaum [2021], El Euch and Rosenbaum [2019]).

En général, les processus dits rugueux ne sont ni Markoviens, ni des semi-martingales. Cela rend l'étude des modèles rugueux particulièrement stimulant aussi bien du point de vue de la simulation des processus que de celui de la valorisation et de la couverture de produits dérivés. A ce sujet, il a été récemment montré dans El Euch and Rosenbaum [2019] que la transformée de Fourier-Laplace d'un modèle rugueux, généralisant celui de Heston, était donnée par une formule semi-explicite impliquant la résolution d'un système d'équations de Riccati fractionnaires. Des techniques d'analyse de Fourier sont alors utilisées pour valoriser des options européennes.

La version rugueuse du modèle de Heston, introduite dans El Euch and Rosenbaum [2019], s'inspire du modèle de Heston classique (voir Heston [1993]), où une convolution avec un noyau fractionnaire est considérée dans la dynamique de la variance. Dans le modèle de Heston rugueux, on a donc le processus de prix de l'actif S solution de l'équation différentielle stochastique:

$$dS_t = S_t \sqrt{V_t} dB_t, \quad \text{pour } t \geq 0, \tag{2}$$

où le processus de variance instantanée V est lui-même solution de:

$$V_t = V_0 + \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)} \left(\lambda \left(\bar{\nu} - V_s \right) ds + \eta \sqrt{V_s} dW_s \right), \quad \text{pour } t \geq 0, \tag{3}$$

où λ, η, V_0 et $\bar{\nu}$ sont des réels strictement positifs, W et B deux mouvements Browniens corrélés. Comme dans le modèle d'origine, λ est le paramètre de vitesse de retour à la moyenne, $\bar{\nu}$ la variance moyenne, V_0 la variance instantanée initiale et η la paramètre de volatilité de la volatilité.

La fonction définie sur $\mathbb{R}^+$ par $t \to t^{\alpha-1}/\Gamma(\alpha)$ est appelée noyau fractionnaire et gouverne la régularité du processus V qui est H-höldérienne avec, $H := \alpha - 1/2$ et $\alpha \in (1/2, 1]$. On note que lorsque $\alpha = 1$, on retrouve le modèle de Heston classique.

On peut exprimer la transformée de Fourier-Laplace des processus du modèle rugueux à l'aide de solutions d'un système d'équations fractionnaires de Riccati. L'approximation numérique de ces solutions est un problème rendu difficile par l'explosion du noyau fractionnaire au voisinage de 0. Plusieurs solutions ont été proposées dans la littérature. Par exemple, dans Callegaro et al. [2021], des schémas de discrétisation hybrides sont utilisés. Dans Abi Jaber and El Euch [2019b] et Abi Jaber [2019a], il est proposé d'approcher les processus rugueux par un processus multi-facteurs qui serait une semi-martingale et bénificierait d'une structure markovienne. Cette approche permet une implémentation efficace du modèle pour valoriser des options européennes.

On a ensuite remarqué que la forme de la transformée de Fourier-Laplace du modèle de Heston rugueux était un cas particulier des structures des transformées de Fourier-Laplace de processus appelés processus affine de Volterra (voir Abi Jaber et al. [2019], Keller-Ressel et al. [2021], Gatheral and Keller-Ressel [2019], Cuchiero and Teichmann [2020]). Ces processus sont solutions d'une équation stochastique de Volterra et sont présentés dans la section suivante.

Équations stochastiques de Volterra

Les équations stochastiques de Volterra sont des équations de convolution d-dimensionnelles de la forme:

$$X_t = u_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad \text{pour } t \geq 0, \tag{4}$$

W est un mouvement Brownien m-dimensionnel. La fonction $u_0 : \mathbb{R}_+ \to \mathbb{R}^d$ est continue, $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ est un noyau de convolution satisfaisant des conditions de régularité et les coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ et $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ satisfont des conditions de régularité et de croissance linéaire.

Si, pour $K = \mathbb{I}_d$, ces équations correspondent au cas classique des équations différentielles stochastiques, pour des noyaux plus généraux, les propriétés de semi-martingale ou de Markovianité des processus peuvent disparaître. L'existence de solutions fortes continues à l'équation

(4) est garantie lorsque b et σ sont Lipschitziennes et l'existence de solutions faibles continues lorsque b et σ ne satisfont qu'une condition de croissance linéaire (voir Abi Jaber et al. [2019], Abi Jaber and El Euch [2019a]). Beaucoup de propriétés de ces équations et de leurs solutions sont rappelées dans le Chapitre 1 de ce manuscrit.

Processus Affines de Volterra

Les modèles classiques de processus affines ont largement été employés en finance, en particulier pour la modélisation de la structure par terme des taux, des risques de crédit et de la valorisation et couverture de produits dérivés (voir Duffie et al. [2003]). Les modèles affines de taux courts les plus connus sont celui de Vasicek (voir Vasicek [1977]), et celui de Cox, Ingersoll, and Ross (CIR) (voir Cox et al. [1985]). Dans le domaine de la valorisation d'options, les modèles affines comme celui de Heston (voir Heston [1993]) sont très utilisés car ils représentent correctement des faits stylisés observés à basse fréquence (voir, par exemple, Drăgulescu and Yakovenko [2002] et Gatheral [2006]).

Commençons par rappeler que les diffusions affines dans un ensemble $E \subseteq \mathbb{R}^d$ sont des solutions faibles de l'équation (4), avec $K = \mathbb{I}_d$, $u_0 = X_0$ et des coefficients b et $a := \sigma\sigma^\top$ affines en les variables d'espace. Un processus X a donc une diffusion affine si il est solution de l équation stochastique différentielle suivante:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad \text{pour } t \geq 0, \tag{5}$$

avec b et $a := \sigma\sigma^\top$ satisfaisant sur E,

$$\begin{aligned} a(x) &= A_0 + x_1 A_1 + \dots x_d A_d, \\ b(x) &= b_0 + x_1 b_1 + \dots x_d b_d, \end{aligned} \tag{6}$$

où les A_i sont des matrices symétriques de taille are $d \times d-$ et les b_i des vecteurs $d-$dimensionnels.

L'attractivité des processus affines est dûe aux possibilités d'obtenir des formules explicites ou relativement faciles à estimer pour beaucoup de leurs fonctionnelles. Par exemple, leurs transformées de Fourier-Laplace sont données par une formule explicite faisant intervenir l'exponentielle d'une fonction affine en espace et dont les coefficients dépendant du temps sont solutions d'une équation de Riccati. Plus précisément, la transformée conditionnelle de Fourier-Laplace d'une solution de l'équation (5) est donnée par:

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \quad \text{pour } t \geq 0, \tag{7}$$

et u la transposée d'un vecteur de nombre complexes, ϕ et ψ solution d'un système d'équations de Riccati que l'on précisera par la suite. Des techniques d'analyse de Fourier sont ensuite à utiliser pour résoudre des problèmes de valorisation, couverture et calibration dans ce cadre (voir, par exemple, Carr and Madan [1999], Fang and Oosterlee [2009]).

Des techniques de variations de la constante donnent une représentation alternative de la transformée de Laplace-Fourier (7), sous la forme suivante

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T|\mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}\left[X_s|\mathcal{F}_t\right])\psi(T-s)^\top ds\right), \tag{8}$$

où ψ est solution de l'équation de Riccati:

$$\psi(t) = u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds,$$

avec $A(u) = (uA_1u^\top, \ldots, uA_du^\top)$ et $B = (b_1 \ldots b_d)$.

Nous nous intéresserons à une classe plus large de processus, contenant les processus affines et constituée des *processus affines de Volterra*. Ces derniers sont définis comme solutions d'équation stochastique de Volterra du type de (4), avec des coefficients a et b satisfaisant (6). En général, ces processus ne sont ni des semi-martingales, ni des processus de Markov, il serait donc surprenant qu'une formule du type de (8) soit vraie. C'est cependant ce qui est établi dans Abi Jaber et al. [2019], avec une fonction ψ solution d'une appelée équation de Riccati-Volterra (ou équation de Riccati fractionnaire si K est un noyau fractionnaire) :

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) K(t-s)ds.$$

L'unicité de solutions faibles d'équations de Volterra du type (4) peut alors être établie comme conséquence de (8). Dans la Section 1.3 nous reviendrons plus en détails sur certains résultats portant sur les processus affines de Volterra et présenterons quelques exemples importants comme les processus de Volterra-Ornstein–Uhlenbeck, Volterra CIR et Volterra Heston models. Le fait que les processus de Volterra ne soient pas des semi-martingales est un autre problème majeur pour leur utilisation car cela rend très difficile de mettre en place des techniques de simulation de ces processus. Nous rappelons maintenant une méthode performante que nous allons utiliser tout au long de notre travail.

Approximations

Les processus, solutions de l'équation stochastique de Volterra (4) peuvent être approchés par un modèle multi-facteurs à volatilité stochastique présentant une structure Markovienne et de semi-martingale. Cette approximation technique repose sur une approximation du noyau K apparaissant dans (4) par une suite de noyaux plus réguliers, qui n'explosent pas en 0.

Considérons des noyaux complètement monotones dont la définition et les propriétés importantes sont rappelées dans la Section 1.4.2. Ce sont des noyaux K de classe $\mathcal{C}^\infty$ sur $(0,\infty)$ tels que $(-1)^j K^{(j)}(t)$ est semi-définie positive pour tout $j \in \mathbb{N}$. Le théorème de Bernstein caractérise l'ensemble des noyaux complètement monotones comme l'ensemble des noyaux qui s'expriment comme transformée de Laplace d'une mesure positive, c'est à dire l'ensemble des noyaux K tels que:

$$K(t) = \int_{\mathbb{R}_+} e^{-xt}\mu(dx), \qquad \text{pour } t \geq 0. \tag{9}$$

Le noyau fractionnaire, défini sur $\mathbb{R}^+$, par $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$, avec $\alpha \in (1/2, 1]$, et utilisé dans la définition de modèle de Heston rugueux (voir (2)-(3)), est un exemple de noyau complètement monotone, associé à la mesure μ donnée par $\mu(dx) = \frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$.

Une approximation de la mesure μ par une somme pondérée de mesures de Dirac est proposée dans Abi Jaber and El Euch [2019b], Carmona et al. [2000], Abi Jaber [2019a], cela mène à approcher le noyau par:

$$K^n(t) = \sum_{i=1}^{n} c_i e^{-x_i t}, \qquad \text{pour } t \geq 0 \text{ et } n \in \mathbb{N}. \tag{10}$$

Cette approximation (10) entraine ensuite l'approximation multi-factorielle de l'équation de (4):

$$\begin{aligned} X_t^n \quad &:= \quad u_0^n(t) + \sum_{i=1}^{n} c_i Y_t^{n,i}, \qquad \text{pour } t \geq 0 \text{ et } n \in \mathbb{N} \qquad (11)\\ dY_t^{n,i} \quad &= \quad \big(-x_i Y_t^{n,i} + b(X_t^n)\big)dt + \sigma(X_t^n)dW_t, \quad i = 1, 2, \ldots, n, \end{aligned}$$

en supposant que $Y_0^{n,i} = 0$ pour tout i $\in \{1, 2, \ldots, n\}$. Le processus X^n défini pour tout entier n, dans (11) peut aussi s'écrire

$$X_t^n = u_0^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s, \qquad t \in [0, T]. \tag{12}$$

On remarque que les processus $Y = (Y_t^{n,i})_{i \leq n}$ définis par les équations (11) et appelés facteurs, sont des semi-martingales Markoviennes. De plus, si l'équation stochastique de Volterra (4) est affine, les processus $Y = (Y_t^{n,i})_{i \leq n}$ sont affines, au sens classique. Abi Jaber and El Euch [2019b, Théorème 3.6] détaille les conditions de convergence en loi de X^n vers X lorsque (K^n, u_0^n) approche (K, u_0).

Cette approximation d'un processus rugueux par des semi-martingales markoviennes a été mise en œuvre et étudiée dans le cadre du modèle de Heston rugueux (2)-(3), (voir Abi Jaber and El Euch [2019b] and Abi Jaber [2019a]). En particulier, des prix d'options Européennes ont été estimés par cette approche et par la résolution d'équations de Riccati plutôt que celle des équations de Riccati-Volterra associées à la transformée de Fourier-Laplace. Cela a permis d'obtenir une méthode plus rapide de valorisation et calibration dans le cadre du modèle de Heston rugueux. Nous montrerons dans le Chapitre 2 que cette approche peut également être bénéfique pour la valorisation d'options Américaines

Valorisation d'options Américaines

Une option Américaine donne le droit à son détenteur d'exercer l'option à n'importe quel moment précédant l'échéance du contrat, par opposition aux options Européennes pour lesquels l'option ne peut être exercer qu'a maturité. Par conséquent, la valorisation d'une option Américaine est un problème d'arrêt optimal, naturellement dépendant de toute la trajectoire du prix du sous-jacent. Plus précisément, si l'on note X le logarithme du processus de prix du sous-jacent, il faudra déterminer le processus de prix de l'option P lorsque le pay-off de l'option est le processus $(f(X_t))_{0 \leq t \leq T}$, avec f définie et continue sur $\mathbb{R}_+$. Sous des hypothèse de viabilité du marché, le processus de valeur de l'option P est donné par le problème d'arrêt optimal suivant :

$$P_t := \operatorname{esssup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[e^{-r(\tau - t)} f(X_\tau) \mid \mathcal{F}_t\right], \qquad \text{pour } 0 \leq t \leq T,$$

où $\mathcal{T}_{t,T}$ est l'ensemble des temps d'arrêt à valeurs dans l'intervalle $[t, T]$. De nombreuses approches ont été développées pour résoudre ce problème dans des cadres markoviens ou lorsque X est une semi-martingale. Cela va de méthodes de Monte-Carlo à des caractérisations du processus valeur comme solution d'équation différentielle rétrograde réfléchie, en passant par l'étude de l'équation d'Hamilton-Jacobi-Bellman associée. Quelque soit l'approche adoptée, elle nécessite une bonne compréhension des lois conditionnelles de X, ce qui est une question stimulante dans un modèle non-Markovien et sans structure de semi-martingale.

Plusieurs travaux récents portent sur la valorisation d'options Américaines dans des modèles à volatilité rugueuse (voir par exemple Goudenège et al. [2020], Bayer et al. [2020a,b], Horvath et al. [2017]). En particulier, des résultats numériques sont obtenus pour le modèle dit de Bergomi rugueux, introduit dans Bayer et al. [2016]. Dans Horvath et al. [2017], il est proposé une approche basée sur l'approximation de Donsker d'un mouvement brownien fractionnaire et sur des techniques d'arbres. Dans Bayer et al. [2020b],les auteurs présentent une méthode de type Monte Carlo avec une optimisation sur le taux d'exercice, toujours pour valoriser numériquement une option Américaine dans un cadre de volatilité rugueuse. Une autre approche est proposée dans Bayer et al. [2020a], où le processus valeur est caractérisé comme solution d'une équation différentielle stochastique rétrograde réfléchie et dépendant de la trajectoire. L'approximation de la solution de cette équation est alors obtenue par des méthodes d'apprentissage profond.

Dans notre travail, dans le Chapitre 2, nous nous intéressons à la valorisation d'option Américaine dans un modèle de Volterra-Heston. Le modèle que nous considérons a été étudié dans Abi Jaber and El Euch [2019a,b], et est une extension des modèles de Heston rugueux (voir (2)-(3)). Le couple de processus composé du prix et de sa variance est un processus affine de Volterra dont nous rappelons la définition et certaines propriétés importantes dans la Section 1.3.

Dans un modèle de Volterra-Heston le processus du logarithme du prix du sous-jacent, X, est donné par:

$$X_t = \int_0^t \left(r - \frac{V_s}{2}\right) \mathrm{d}s + \int_0^t \sqrt{V_t} dB_s, \quad \text{pour } t \geq 0, \tag{13}$$

et celui de la variance V est un processus solution de l'équation stochastique de Volterra suivante:

$$V_t = v_0(t) - \lambda \int_0^t K(t-s) V_s \,\mathrm{d}s + \eta \int_0^t K(t-s)\sqrt{V_s}\,\mathrm{d}W_s, \quad \text{pour } t \geq 0, \tag{14}$$

où r est le taux d'intérêt sans risque, $B = \rho W + \sqrt{1-\rho^2} W^{\perp}$, avec $(W, W^{\perp})$ mouvement Brownien de dimension deux défini sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $\rho \in [-1, 1]$, λ et η réels positifs quantifiant respectivement la vitesse de retour ̀la moyenne et la volatilité du processus de variance. Le noyau de convolution K est une fonction de $L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ et $v_0 : \mathbb{R}_+ \to \mathbb{R}$ est une fonction déterministe continue et satisfaisant des conditions de régularité supplémentaires.

Nous commencerons par approcher le processus (X, V) par la procédure décrite dans la section précédente. Il s'agit d'approcher le noyau complétement monotone K par une suite de noyau $(K_n)_{n\geq 0}$ construits comme somme pondérée de mesure de Dirac (voir Carmona et al. [2000], Harms and Stefanovits [2019]). Le processus (X^n, V^n) approximant (X, V) est alors Markovien et a une structure de semi-martingale, tout en restant un processus affine. En effet, il est solution de l'équation stochastique de Volterra suivante:

$$\begin{aligned} \mathrm{d}X_t^n &= \left(r - \frac{V_t^n}{2}\right)\mathrm{d}t + \sqrt{V_t^n} dB_t^n, \\ V_t^n &= v_0^n(t) - \lambda \int_0^t K^n(t-s) V_s^n \,\mathrm{d}s + \eta \int_0^t K^n(t-s)\sqrt{V_s^n}\,\mathrm{d}W_s^n. \end{aligned} \tag{15}$$

On sait que cette approximation multi-factorielle (X^n, V^n) converge en loi vers un processus de Volterra Heston solution de (13)-(14) dès que (K^n, u_0^n) tend vers (K, u_0) , dans un sens que l'on précisera (voir Abi Jaber and El Euch [2019b]).

Lorsque l'on travaille avec les processus approchés, le cadre correspond à un modèle Markovien de grande dimension avec une structure de semi-martingale affine. De nouvelles techniques pour aborder ce cadre et valoriser numériquement une option Américaine, ont été récemment développées (voir, par exemple, Goudenège et al. [2020], Lapeyre and Lelong [2019]). Dans Goudenège et al. [2020], des méthodes efficaces, basées sur du machine learning, permettent d'évaluer des prix d'options Américaines dans des modèles Markoviens de grande dimension. Une adaptation de ces méthodes est également proposée pour traiter les cas non Markovien du modèle de Bergomi rugueux. Les auteurs de Lapeyre and Lelong [2019], utilisent des réseaux de neurones pour valoriser des options Bermudéennes. Par opposition aux techniques de régression usuellement appliquées, ils emploient des réseaux de neurones pour approcher les espérances conditionnelles apparaissant dans l'équation de programmation dynamique dont est solution le processus de prix de l'option Bermudéenne. Cela débouche sur une implémentation performante du bien connu algorithme de Longstaff-Schwartz.

Il nous semble important de mentionner ici que notre étude ne porte pas sur la recherche de méthodes efficaces pour implémenter le modèle multi-factoriel mais sur la convergence des prix évalués dans ce modèle vers ceux obtenus dans le modèle limite de Volterra.

Le résultat principal de ce chapitre dans le Théorème 2.1.7 est donc que le prix d'une option

Américaine approché dans le modèle multi-factoriel (15), et noté P_0^n, converge vers le prix P_0 de l'option dans le modèle de Volterra Heston original (13)-(14) sous des conditions de régularité de K^n et u_0^n.

Dans la première partie de notre étude, nous établissons la convergence du prix d'une option Bermudéenne dans le modèle multi-factoriel, (X^n, V^n), vers le prix d'une Bermudéenne dans le modèle limite de Volterra Heston. Ce résultat repose sur la structure affine des processus du modèle de Volterra Heston qui nous permet de prouver la convergence des transformées de Fourier-Laplace conditionnelles.

Un processus auxiliaire va jouer un rôle crucial dans ce dernier résultat, il s'agit d'un processus associé à V que nous qualifierons de forward ajusté. Introduit et étudié dans Abi Jaber and El Euch [2019a], il caractérise la structure Markovienne de dimension infinie du modèle de Volterra Heston. Le processus forward ajusté de V est donné par:

$$v_t(\xi) \quad = \quad \mathbb{E}\left[V_{t+\xi} + \lambda \int_0^\xi K(\xi - s) V_{t+s} ds \middle| \mathcal{F}_t\right], \qquad \xi \geq 0.$$

Il est à noter que $v_t(0) = V_t$. Les propriétés de ce processus nous apporterons une meilleure compréhension des lois conditionnelles du modèle.

Une fois la convergence des options Bermudéennes établie, nous montrons la convergence du prix d'une option Américaine dans le modèle multi-factoriel (X^n, V^n) vers le prix d'une Bermudéenne dans le modèle limite de Volterra Heston. Ce résultat est une conséquence d'arguments classiques (voir Lamberton and Pagès [1990]).

Nous terminons ce chapitre par une présentation de résultats numériques. Nous exploitons la propriété de surmartingale du processus (X^n, V^n) pour le simuler par un schéma implicite-explicite, puis estimons le prix d'une option Américaine à l'aide de l'algorithme de Longstaff Schwartz (voir Longstaff and Schwartz [2001]).

Moments

Dans le Chapitre 3, nous nous intéressons à une classe plus large de processus, contenant les processus affines de Volterra, appelée processus polynômiaux de Volterra. Nous commençons par une brève étude des diffusions polynômiales. Une telle diffusion est une solution faible de l'équation stochastique de Volterra (5), avec $K = 1$, telle que le coefficient de dérive b soit affine et le coefficient de diffusion $\sigma\sigma^\top$ soit quadratique en les variables d'espace. Les diffusions polynômiales sont utilisées dans beaucoup d'applications à la finance comme les modèles de structure par terme des taux de swap sur la variance (voir Filipović et al. [2016]) et des taux d'intérêts (voir Filipović et al. [2017]), les modèles de gestion du risque de crédit (voir Ackerer and Filipović [2020]), ou encore des modèles utilisant des techniques de réduction de la variance pour la valorisation d'options (voir Cuchiero et al. [2012]).

Une propriété importante de ces diffusions polynômiales est que leurs moments conditionnels sont donnés par une formule explicite (voir Cuchiero et al. [2012], Filipović and Larsson [2016]). En effet, ces moments conditionnels peuvent être évalués par le calcul de l'exponentiel d'une matrice. On peut montrer en particulier que si X est une diffusion polynômiale, alors ses moments sont des fonctions polynômiales de l'état initial X_0. L'efficacité et la relative facilité de calcul des moments permet, par exemple, de développer une méthode performante de valorisation d'option, basée sur des techniques d'approximation de densité (voir Ackerer et al. [2018], Filipović et al. [2013], Heston and Rossi [2017]).

Notre but, dans le Chapitre 3 est de retrouver certaines des propriétés essentielles des moments de diffusions polynômiales dans un cadre d'équations polynômiales de Volterra. Nous considèrons donc un processus polynômial de Volterra, solution faible de l'équation (4) avec b affine et a quadratique. On remarque que les processus affines de Volterra sont également des processus polynômiaux de Volterra.

Nos premiers résultats concernent le cas particulier des processus affines de Volterra. Le Théorème 3.3.3 établit que les moments d'ordre p d'une diffusion affine de Volterra est un polynôme de degré inférieur à p appliqué à la valeur initiale du processus.

Nous présentons aussi des formules générales pour calculer les moments conditionnels des processus affine de Volterra en dimension 1 dans les Sections 3.3.2 et 3.3.3.

Dans le Théorème 3.4.1 nous considèrons une suite $X^n = (X^n_t)_{t\leq T}$ de solutions faibles continues de l'équation stochastique de Volterra (4) qui converge faiblement vers X solution de cette même équation (4). Sous des conditions d'existence de bornes uniformes pour X^n, on obtient la convergence des moments de X^n vers ceux de X:

$$\mathbb{E}[P(X^n)] \to \mathbb{E}[P(X)], \quad \text{pour tout polynôme } P \text{ sur } \mathbb{R}^d.$$

Ce dernier Théorème nous donne une méthode efficace de calcul des moments des processus polynômiaux de Volterra. En appliquant les techniques d'approximation décrites dans la Section 1.4.2, nous pouvons construire un processus multi-factoriel X^n qui approche X. Une remarque importante est que les facteurs apparaissant dans X^n sont des diffusions polynômiales classiques. Ainsi, nous pouvons évaluer leurs moments en calculant des exponentielles de matrices. Nous terminerons ce chapitre avec quelques exemples numériques.

Estimation de paramètres

L'étude statistique des processus de diffusion dirigés par un mouvement Brownien standard a une longue histoire. Le probème d'estimation du terme de dérive d'un processus d'Itô qui peut être observé en continu ou en temps discret a été largement étudié (voir par exemple Kutoyants [2004], Liptser and Shiryaev [2001], Basawa and Scott [1983]).

Parmi les processus d'Itô les plus étudiés on trouve le processus d'Ornstein-Uhlenbeck (voir [Uhlenbeck and Ornstein, 1930]), processus Markovien, Gaussien , continu, solution de l'équation différentielle stochastique suivante:

$$X_t = \theta X_t dt + \sigma dW_t, \quad 0 \leq t \leq T, \tag{16}$$

où $\theta \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ est le paramètre de volatilité et W un mouvement brownien standard.

L'estimation paramètrique de θ, vitesse de retour à la moyenne du processus d'Ornstein-Uhlenbeck est motivée par le besoin pratique d'améliorer la modélisation de la volatilité dans les modèles financiers dits à volatilité stochastique. En effet, des processus avec retour à la moyenne modélisent la variance des prix dans ce type de modèles.

Dans le cas des processus d'Ornstein-Uhlenbeck, le paramètre de volatilité σ peut être estimé grâce à la variation quadratique de la série temporelle observée. La question de l'estimation de

θ est beaucoup plus délicate. Il y a principalement deux estimateurs populaires: l'estimateur de maximum de vraisemblance et celui des moindres carrés. Les deux estimateurs coincident et sont fortement consistant dans le cas des processus de d'Ornstein-Uhlenbeck. L'estimateur de maximum de vraisemblance de θ est un processus, noté $\hat{\theta}$ qui, après application du Théorème de Girsanov (voir Liptser and Shiryaev [2001]), vaut:

$$\hat{\theta}_t = \frac{\int_0^t X_s dX_s}{\int_0^t X_s^2 ds}, \quad \text{pour } t \geq 0. \tag{17}$$

En finance, les praticiens se sont intéressés à l'étude de modèles de volatilité rendant compte de phénomènes de mémoire longue. Cette caractéristique peut être obtenue en remplaçant le mouvement Brownien de l'équation (16) par sa version fractionnaire. Le processus obtenu est alors appelé processus d'Ornstein-Uhlenbeck fractionnaire et est l'unique processus gaussien solution de l'équation stochastique de Volterra:

$$X_t = \theta \int_0^t X_s ds + \sigma W_t^H, \quad 0 \leq t \leq T, \tag{18}$$

où W^H est un mouvement brownien fractionnaire (1) de paramètre de Hurst $H \in (0, 1)$. Le paramètre de Hurst H caractérise la régularité höldérienne des trajectoires de X. Pour le mouvement Brownien standard, $H = 1/2$ et ses accroissements sont indépendants. Si $H > 1/2$ les accroissements sont corrélés positivement et les trajectoires plus régulières. Lorsque $H < 1/2$, les accroissements sont négativement corrélés et les trajectoires sont moins régulières. L'estimation du paramètre de dérive d'un processus d'Ornstein-Uhlenbeck fractionnaire est un problème rendu difficile car le mouvement Brownien fractionnaire n'est ni Markovien, ni une semi-martingale lorsque $H \neq 1/2$.

Le comportement dit à mémoire longue des processus de variance est satisfait lorsque $H > 1/2$. Dans ce cas, il ya eu plusieurs travaux étudiant la consistance forte des estimateurs de vraisemblance et des moindres carrés pour des observations du processus continues (voir Kleptsyna and Le Breton [2002], Hu and Nualart [2010], Belfadli et al. [2011]) et discrètes (voir Cénac and Es-Sebaiy [2015], Es-Sebaiy and Ndiaye [2014], Hu and Song [2013], Xiao et al. [2011], Bishwal [2011]).

Cependant, comme mentionné au début de cette introduction, de récentes études ont démontré que la volatilité des actifs financiers était rugueuse, c'est à dire que le bon choix de paramètre de Hurst est $H < 1/2$. Dans ce cas, il y a également eu des études de consistance des deux estimateurs principaux pour des observations continues (voir par exemple Hu et al. [2019], Tudor and Viens [2007], El Machkouri et al. [2016]) et discrètes (voir par exemple Tudor and Viens [2007], Kubilius et al. [2015], Es-Sebaiy et al. [2019]) .

Dans le Chapitre 4, nous étudions le problème statistique d'estimation du paramètre de vitesse de retour à la moyenne d'un processus d'Ornstein–Uhlenbeck dit de Volterra. A notre connaissance, c'est le premier essai de résolution de ce problème. Le processus d'Ornstein - Uhlenbeck de Volterra est un cas particulier des processus affines de Volterra et il est défini comme l'unique solution forte de l'équation stochastique de Volterra suivante:

$$X_t = \int_0^t K(t-s)(\theta X_s ds + dW_s) \quad t \geq 0, \tag{19}$$

où $\theta \in \mathbb{R}$, est un paramètre inconnu Le noyau K que nous considèrerons est le noyau fractionnaire défini pour $t > 0$ par $K(t) = t^{\alpha-1}/\Gamma(\alpha), \quad \alpha \in (1/2, 1]$. $\alpha = H + 1/2$ gouverne la régularité des trajectoires du processus. Notons que lorsque $K = 1$ et $\alpha = 1$, nous retrouvons le processus classique d'Ornstein–Uhlenbeck.

L'estimation de ce pramètre θ pour un processus d'Ornstein-Uhlenbeck de Volterra est un problème stimulant car, comme pour le processus d'Ornstein-Uhlenbeck fractionnaire, le processus n'est ni markovien ni une semi-martingale. Dans le cadre Volterra, le noyau agit aussi sur le paramètre que l'on souhaite estimer multipliant les difficultés dans notre études statistique.

Notre but est donc d'estimer θ dans (19). Nous commençons par étudier cette estimation lorsque l'on a accès à des observations continues X. L'estimateur du maximum de vraisemblance de θ est le processus $\hat{\theta}$ donné par:

$$\hat{\theta}_t = \frac{\int_0^t X_s dZ_s}{\int_0^t X_s^2 ds}, \quad \text{pour } t \geq 0, \tag{20}$$

où Z est un processus auxiliaire que nous utiliserons dans la construction et l'étude de l'estimateur, notamment car ce processus, contrairement à X, est une semi-martingale.

Notre résultat principal dans le Théorème 4.2.1 est que l'estimateur $\hat{\theta}$ défini dans (20) est fortement consistant pour tout $\theta > 0$, c'est à dire que

$$\lim_{t \to \infty} \hat{\theta}_t = \theta, \quad \text{a.s.}$$

Les observations d'un processus sont en général des échantillons discrets de valeurs. Par conséquent, pour des questions pratiques, l'étude du cas avec observations discrètes a un intérêt significatif. Nous nous intéressons donc au problème d'estimation de θ pour un processus d'Ornstein-Uhlenbeck de Volterra que l'on observe de façon discrète seulement. Nous suivrons l'approche développée dans Kubilius et al. [2015] pour les processus d'Ornstein-Uhlenbeck fractionnaires. Nous supposerons que le processus X est observé aux dates $\{t_k, n \geq 1, 0 \leq k \leq a_n\}$ avec a suite d'entier tendant vers l'infini et considèrerons l'estimateur de vraisemblance sous sa version discrétisée, défini, pour $n \in \mathbb{N}$, par

$$\bar{\theta}_n(a) = \frac{\sum_{k=0}^{a_n-1} X_{t_k} \Delta Z_{t_k}}{\frac{1}{n} \sum_{k=0}^{a_n-1} X_{t_k}^2}, \tag{21}$$

où $\Delta Z_{t_k} = Z_{t_{k+1}} - Z_{t_k}$.

Notre résultat principal dans ce contexte dans le Théorème 4.3.1 est que pour $\theta \in \mathbb{R}$, l'estimateur $\bar{\theta}(a)$ défini dans (21) est consistant lorsque la suite a rempli certaines conditions limite à l'infini.

La démonstration de ce résultat repose principalement sur le fait que X, défini dans (19), est un processus gaussien dont on a une expression explicite (voir Abi Jaber et al. [2019]) :

$$X_t = \int_0^t K_\theta(t-s)\sigma dW_s, \quad \text{pour } t \geq 0,$$

où le noyau K_θ est défini sur $\mathbb{R}^+$ par $K_\theta(t) = t^{\alpha-1} E_{\alpha,\alpha}(\theta t^\alpha)$ avec $E_{\alpha,\alpha}$ fonction de Mittag-Leffler définie sur $\mathbb{C}$ par:

$$E_{\alpha,\alpha}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \alpha)}.$$

Ainsi, la démonstration de consistance de l'estimateur requiert une bonne compréhension des propriétés asymptotiques des fonctions de Mittag-Leffler car cela nous permettra d'estimer la variance de X_t pour t grand.

Nous concluons ce chapitre par une présentation de résultats numériques et des méthodes utilisées pour les obtenir. Il est à noter que la simulation d'un processus d'Ornstein-Uhlenbeck de Volterra fait appel aux techniques d'approximation multi-factorielle, présentées dans la Section 1.4.2.

Introduction (English version)

In the realm of mathematical models to reproduce the dynamics in financial markets, models for pricing derivatives have been of great interest. The most popular model for pricing derivatives is the Black and Scholes model Black and Scholes [1973]. In this model the stock price S follows an Itô diffusion:

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

where μ is the drift parameter, B represents a standard Brownian motion and σ denotes the volatility parameter. Even though the pricing and hedging of European options can be efficiently implemented, this model is not in accordance with the reality in financial markets. Among other inconsistencies, it can be observed that the volatility of an asset is not constant. To overcome these limitations a new class of models have appeared, namely stochastic volatility models. In these models the volatility V follows a stochastic differential by its own:

$$dV_t = \alpha_{V,t} dt + \beta_{V,t} \sigma dW_t,$$

where $\alpha_{V,t}$ and $\beta_{V,t}$ are some functions of V and W is a Brownian motion correlated with B. These models in general have a semimartingale continuous Markovian structure. Stochastic volatility models, such as the Heston model Heston [1993], reproduce correctly the term structure of at-the-money skew for long maturities but they fail to explain the behavior for very short maturities.

In fact, let $\sigma_{BS}(k, \tau)$ [2] be the implied volatility of an option with log-moyeness k and time to maturity τ. The term structure of at-the-money (ATM) skew is given by:

$$\xi(\tau) = \left| \frac{\partial \sigma_{BS}(k, \tau)}{\partial k} \right|_{k=0}$$

[2] $\sigma_{BS}(k, \tau)$ is the volatility parameter in the Black and Scholes option pricing model, that gives the market price of an option.

Empirically studies have shown a power-law decay of the ATM skew of option prices, with an explosion when τ is small. In Fukasawa [2011], it is shown that at least for small τ the term structure of a stochastic volatility model, where the volatility is driven by a fractional Brownian motion with Hurst parameter H, is given by $\xi(\tau) = \tau^{H-1/2}$. The use of the fractional Brownian motion was initiated by Comte and Renault [1998] who proposed to model log-volatility using the fractional version W^H of the classical Brownian motion:

$$W_t^H = \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^{0} \left((t-s)^{H-1/2} - (-s)^{H-1/2} \right) dW_s + \int_0^t (t-s)^{H-1/2} dW_s \right), \qquad (22)$$

where $H \in (0,1)$ is called the Hurst parameter and governs the regularity of the trajectories of W^H. The sample paths of W^H are locally Hölder continuous of any order strictly less than H. Comte and Renault [1998] chose the Hurst parameter $H > 1/2$ to ensure long memory. However, Gatheral et al. [2018] demonstrates that the log-volatility behaves essentially as a fractional Brownian motion with Hurst parameter H of order 0.1.

This observation has raised a growing interest in the so-called rough volatility models because they can reproduce the power-law decay of the ATM skew [Alòs et al., 2007, Fukasawa, 2011, Bayer et al., 2016, Garnier and Solna, 2017, Forde and Zhang, 2017, Fukasawa, 2017, Guennoun et al., 2018, Bayer et al., 2019, Euch et al., 2019, Fukasawa, 2021]. Rough volatility models are stochastic volatility models whose trajectories are continuous but rougher than the paths of a Brownian motion in terms of their Hölder regularity. Specifically, the trajectories are considered as rough when the Hölder regularity is less than 1/2.

Moreover, rough models seem remarkably consistent with financial time series data. Statistical evidence, under multiple time scales and across many markets, supports the fact that the time series of realized volatility oscillates more rapidly than Brownian motion [Bennedsen et al., 2021, Gatheral et al., 2018, Livieri et al., 2018, Fukasawa et al., 2019]

In addition, these discoveries are supported by micro-structural considerations because rough volatility models appear naturally as scaling limits of micro-structural pricing models with self-exciting features driven by Hawkes processes [Jaisson and Rosenbaum, 2016, El Euch et al., 2018, Dandapani et al., 2019, Tomas and Rosenbaum, 2021, El Euch and Rosenbaum, 2019].

In general, these models are neither Markov non semi-martingales, making challenging their theoretical study and in practice, the pricing of derivatives. Nevertheless, in El Euch and Rosenbaum [2019] it has been shown that the Fourier-Laplace transform in the rough Heston model has a semi-explicit formula modulo the resolution of a system of fractional Riccati equations. Fourier-based techniques can be then used to price European options.

The rough Heston model, first introduced by El Euch and Rosenbaum [2019], is the rough version of the well-known Heston model Heston [1993], where a convolution with a fractional kernel is considered in the variance dynamics. In this model, the asset price S satisfies:

$$dS_t = S_t \sqrt{V_t} dB_t, \tag{23}$$

and the instantaneous variance process V is given by:

$$V_t = V_0 + \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)} \left(\lambda \left(\bar{\nu} - V_s \right) ds + \eta \sqrt{V_s} dW_s \right), \tag{24}$$

where λ, η, V_0 and $\bar{\nu}$ are positive and W and B are two Brownian motion with correlation $\rho \in [-1, 1]$. The parameters in this model play the same role as in the Heston model. That is, λ is the mean reversion rate, $\bar{\nu}$ is the long run variance, V_0 the initial variance and ν is the parameter representing the volatility of volatility. $K(t) = t^{\alpha-1}/\Gamma(\alpha)$, $\alpha = H + 1/2$ is called the fractional kernel, where $\alpha \in (1/2, 1]$ governs the smoothness of the volatility. Notice that when $\alpha = 1$, we recover the Heston model.

The numerical resolution of the fractional Riccati equations appearing in the Fourier-Laplace transform of the rough Heston model can be challenging. This is due to the exploding character of the fractional kernel. There has been some research in order to find a way to make the implementation of this model efficient. For instance in Callegaro et al. [2021], fast hybrid schemes are used to solve the fractional Riccati equations. In Abi Jaber and El Euch [2019b] and Abi Jaber [2019a], they build a tractable multi-factor semimartingale approximation of the rough Heston model which enjoys of a Markovian structure. This has lead to an efficient implementation of this model to price European options.

The form of the Fourier-Laplace transform in the rough Heston model is a particular case of the structure of the Fourier-Laplace transforms of the so-called affine Volterra processes [Abi Jaber et al., 2019, Keller-Ressel et al., 2021, Gatheral and Keller-Ressel, 2019, Cuchiero and Teichmann, 2020]. These processes satisfy stochastic Volterra equations study in the next section.

Stochastic Volterra equations

Stochastic Volterra equations are d-dimensional stochastic convolution equations of the form:

$$X_t = u_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{25}$$

where W is a m-dimensional Brownian motion. The function $u_0 : \mathbb{R}_+ \to \mathbb{R}^d$ is continuous, $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ is a convolution kernel that satisfies some regularity conditions and the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous and satisfy a linear growth condition.

Notice that the case $K = \mathbb{I}_d$ and $u_0 = X_0$ corresponds to classical stochastic differential equations. For more general kernels, stochastic Volterra equations do not fall in general in the semi-martingale and Markov setting. The existence of continuous strong solutions to (25) is guaranteed when b and σ are Lipschitz continuous and the existence of continuous weak solution when they satisfy a linear growth condition (see Abi Jaber et al. [2019], Abi Jaber and El Euch [2019a]). Some known properties of these equations are stated in Chapter 1.

Affine Volterra Difussions

Classical affine models have been widely employed in finance, specially in term structure models, credit risk and option pricing (see Duffie et al. [2003]). Well known interest rate models as Vasicek [Vasicek, 1977], and Cox, Ingersoll, and Ross (CIR) [Cox et al., 1985], are examples of affine processes. In the realm of asset pricing modelling, affine models such as the Heston model [Heston, 1993], are popular because they have properties that suit low frequency observations in financial markets (see, for example Drăgulescu and Yakovenko [2002] and Gatheral [2006]).

Affine diffusions in $E \subseteq \mathbb{R}^d$ are weak solutions of (25), with $K = \mathbb{I}_d$ and $u_0 = X_0$. Hence, they satisfy the following stochastic differential equations:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{26}$$

where $b(x)$ and $a(x) = \sigma(x)\sigma(x)^\top$ are such that their coefficients are affine in x,

$$a(x) = A_0 + x_1A_1 + \dots x_dA_d, \tag{27}$$

$$b(x) = b_0 + x_1b_1 + \dots x_db_d, \tag{28}$$

where A_i are $d-$dimensional symmetric matrices and b_i are $d-$dimensional vectors.

The attractiveness of affine processes arises from the fact that they exhibit a high degree of tractability. Their Fourier-Laplace transform has a simple form. It can be explicitly computed as an exponential function which is affine in the space variable and whose coefficients solve a system of Riccati equations. More precisely, the Fourier-Laplace of (26) has the following form:

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \tag{29}$$

where u is a row complex-vector and, ϕ and ψ solve a system of Riccati equations in terms of a and b. Fourier transformation techniques are then used to solve pricing, calibration or hedging problems (see, for example, Carr and Madan [1999], Fang and Oosterlee [2009]).

The variation of constants formula applied to X and ψ, gives the following alternative form to write the Laplace-Fourier transform (29),

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T|\mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}\left[X_s|\mathcal{F}_t\right])\psi(T-s)^\top ds\right), \tag{30}$$

where the row vector function ψ satisfies the Riccati-equation:

$$\psi(t) = u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds,$$

with $A(u) = (uA_1u^\top, \ldots, uA_du^\top)$ and $B = (b_1 \ldots b_d)$.

Affine diffusions are a particular examples of more general processes called affine Volterra processes. Affine Volterra processes are solutions of the stochastic Volterra equation (25), where a and b satisfy (27). In general, these processes do not fall in the semi-martingale and Markov setting. Consequently, one can not expect that a Markovian formula like (30) holds. However, in Abi Jaber et al. [2019], it has been shown that the Fourier-Laplace transform of affine Volterra processes also satisfy (30), where ψ solves the Riccati-Volterra equation:

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) K(t-s)ds. \tag{31}$$

Weak uniqueness for solutions of (25) is then establish as a consequence of (30). In Section 1.3 we present some known results of the affine Volterra process and show some examples corresponding to the Volterra Ornstein–Uhlenbeck, Volterra square-root (Volterra CIR process), and Volterra Heston models. The fact that the Volterra processes are not semi-martingales is another major problem for their use because it makes it difficult to implement simulation techniques. We now recall an efficient method that we will use throughout our work.

Approximation

Stochastic Volterra equations in (25) can be approximated by tractable multifactor stochastic volatility models which enjoy of a semimartingale Markovian structure. This approximation technique relies in the approximation of the kernel K in (25) by a sequence of smoothed ones.

We consider completely monotone kernels and study them in Section 1.4.2. That is, kernels K infinitely many times differentiable on $(0,\infty)$ such that $(-1)^j K^{(j)}(t)$ is positive semi-definite for all $j \in \mathbb{N}$. Bernstein's theorem states that K is completely monotone if and only if K can be expressed as the Laplace transform of a positive $d \times d$-measure μ:

$$K(t) = \int_{\mathbb{R}_+} e^{-xt}\mu(dx). \tag{32}$$

The fractional kernel $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$, with $\alpha \in (1/2, 1]$, used in the rough Heston model (23)-(24), is an example of a completely monotone kernel and its associated measure μ is given by $\mu(dx) = \frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$.

An approximation of the measure μ in (32) with a weighted sum of Dirac measures, as studied in Abi Jaber and El Euch [2019b], Carmona et al. [2000], Abi Jaber [2019a], yields the following smoothed approximation of the kernel:

$$K^n(t) = \sum_{i=1}^n c_i e^{-x_i t}. \tag{33}$$

The representation (33) yields the following factor-representation for the Volterra equation (25):

$$\begin{aligned} X_t^n &= u_0^n(t) + \sum_{i=1}^n c_i Y_t^{n,i}, \\ dY_t^{n,i} &= \big(-x_i Y_t^{n,i} + b(X_t^n)\big)dt + \sigma(X_t^n)dW_t, \quad i = 1, 2, \dots, n, \end{aligned} \tag{34}$$

with initial value $Y_0^{n,i} = 0$ for all i $\in \{1, 2, \dots, n\}$. The approximation X^n in (34) can be written as:

$$X_t^n = u_0^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s, \qquad t \in [0,T]. \tag{35}$$

We remark that the processes $Y = (Y_t^{n,i})_{i\leq n}$ in (34), called factors, are Markovian semimartingales. Hence, the process X can have a Markovian-semimartingale approximation X^n given by a weighted sum of n factors. Moreover, if the stochastic Volterra equation (25) is affine, then the processes $Y = (Y_t^{n,i})_{i\leq n}$ are affine in the classical sense. Abi Jaber and El Euch [2019b, Theorem 3.6] gives the conditions for the convergence in law of X^n to X when (K^n, u_0^n) converges to (K, u_0).

In the rough Heston model (23)-(24), this approximation has been used to approximate the rough model by Markovian semimartingales (see Abi Jaber and El Euch [2019b] and Abi Jaber [2019a]). In particular, they approximate European option prices solving classical Riccati equations instead of the Volterra Riccati equations associated with the characteristic function. This leads to a faster pricing and calibration of European options. In Chapter 2 we will show that this approximation is useful to price American options.

Pricing American options

American options allow holders to exercise the option rights at any time up to the option's expiration date, opposed to European options that can be exercised only at maturity. Therefore, pricing American options is a path-dependent problem. More precisely, denote by X the asset's log return, we are interested in determining the value process P of options with pay-off $(f(X_t))_{0\leq t\leq T}$, where f is a continuous function defined on $\mathbb{R}_+$. The value process P is given by:

$$P_t := \operatorname{esssup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\left[e^{-r(\tau-t)} f(X_\tau) \mid \mathcal{F}_t\right], \quad \text{for } 0\leq t\leq T,$$

where $\mathcal{T}_{t,T}$ denotes the set of all stopping times taking values in $[t, T]$. Many approaches have been developed to solve this problem in Markovian frameworks or when X is a semimartingale. This ranges from Monte-Carlo methods to characterizations of the value process as a solution of a reflected backward stochastic differential equation, passing through the study of the associated Hamilton-Jacobi-Bellman equation. Whatever approach is taken, it requires a good understanding of the conditional laws of X, which is a challenging question in a non-Markovian model and without a semi-martingale structure.

There has been some studies regarding the pricing of American options under rough volatility (see par example Goudenège et al. [2020], Bayer et al. [2020a,b], Horvath et al. [2017]). Specially, there are some numerical results for the rough Bergomi model introduced in Bayer et al.

[2016]. Horvath et al. [2017], propose an approach based on Donsker's approximation for fractional Brownian motion and develop three-base techniques. In Bayer et al. [2020b], they present a method based on Monte Carlo simulation and exercise rate optimization for the numerical pricing of American options. In Bayer et al. [2020a], the authors study the option pricing problem for rough volatility models. In particular, they approximate American option prices using backward stochastic differential equations (BSPDEs) satisfied by the value of the problem. The approximations of these BSPDEs is done by using a deep learning-based method.

In the present work, in Chapter 2, we tackle the problem of pricing American options in the Volterra Heston model. We consider the Volterra Heston model as studied in Abi Jaber and El Euch [2019a,b], which includes the rough Heston model (23)-(24) as a particular case. Moreover, this model belongs to the class of affine Volterra process study in Section 1.3.

In the Volterra Heston model the the asset's log return X is given by:

$$X_t = \int_0^t \left(r - \frac{V_s}{2} \right) \mathrm{d}s + \int_0^t \sqrt{V_t} dB_s, \tag{36}$$

and the instantaneous variance process V is a Volterra square root process that solves the stochastic Volterra equation:

$$V_t = v_0(t) - \lambda \int_0^t K(t-s) V_s \, \mathrm{d}s + \eta \int_0^t K(t-s) \sqrt{V_s} \, \mathrm{d}W_s, \quad t \geq 0, \tag{37}$$

where r is the risk-free rate, $B = \rho W + \sqrt{1-\rho^2} W^{\perp}$, with $(W, W^{\perp})$ a two-dimensional Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\rho \in [-1, 1]$, λ and η are positive and represent the mean reversion speed and the volatility of volatility, respectively. The convolution ke $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ is a convolution kernel and $v_0 : \mathbb{R}_+ \to \mathbb{R}$ is a continuous function satisfying some regularity conditions.

We use the approximation procedure described before to find an approximation (X^n, V^n) which enjoys of Markovian-semimartingale structure, together with the affine diffusion property. We approximate the completely monotone kernel K by a sequence of smoothed kernels K^n given by a weighted sum of Dirac measures, idea initially inspired by Carmona et al. [2000], Harms and Stefanovits [2019]. This in turn leads to an approximation of the Volterra Heston model in (36)-(37), given by:

$$\begin{aligned} \mathrm{d}X_t^n &= \left(r - \frac{V_t^n}{2}\right)\mathrm{d}t + \sqrt{V_t^n} dB_t^n, \\ V_t^n &= v_0^n(t) - \lambda \int_0^t K^n(t-s) V_s^n \,\mathrm{d}s + \eta \int_0^t K^n(t-s)\sqrt{V_s^n}\,\mathrm{d}W_s^n. \end{aligned} \tag{38}$$

This multi-factor approximation (X^n, V^n) in (38) converges in law to the Volterra Heston model (36)-(37) whenever (K^n, u_0^n) converges to (K, u_0) [Abi Jaber and El Euch, 2019b].

This multi-factor approximation is a high-dimensional model enjoying of affine semimartingale Markovian structure. There are some new techniques (see for example Goudenège et al. [2020], Lapeyre and Lelong [2019]) that allow to perform pricing of American options under this kind of models. In Goudenège et al. [2020], they present efficient techniques based on machine learning to compute the price of American options in high-dimensional Markovian models. They also adapt their methods for pricing American options in the non-Markovian case of the rough Bergomi model. In Lapeyre and Lelong [2019], the authors use neural networks to price Bermudan options. Specifically, opposed to the classical regression techniques they employ neural networks to approximate the conditional expectations appearing in the computation of the continuation value. This leads to an efficient implementation of the well-known Longstaff and Schwartz algorithm to price Bermudan option. This is specially handy when working with high dimensional problems.

It is important to mention that we do not focus on the study of efficient ways to implement the multi-factor model but on the convergence of the prices in this model towards the prices in the limiting Volterra model.

Our main result in Theorem 2.1.7 shows that the approximated American option prices denoted by P_0^n in the multi-factor model (38) converge to the prices P_0 in the original Volterra Heston model (36)-(37) under some regularity conditions over K^n and u_0^n.

In the first part of this theorem we concentrate on the convergence of prices of Bermudan options in the multi-factor model (X^n, V^n) towards the prices in the limiting Volterra model. In order to prove the convergence of prices in Bermudan options, we make use the affine structure of the Volterra Heston model, showing first the convergence of the conditional Fourier-Laplace transforms.

A key process to prove the convergence of Bermudan options is the adjusted forward process. This process has been studied in Abi Jaber and El Euch [2019a] to characterize the Markovian structure of the Volterra Heston model. The adjusted forward process of V is given by:

$$v_t(\xi) \quad = \quad \mathbb{E}\left[V_{t+\xi} + \lambda \int_0^{\xi} K(\xi - s)V_{t+s}ds \middle| \mathcal{F}_t\right], \qquad \xi \geq 0.$$

Notice that $v_t(0) = V_t$. This process allows to have a better understanding of the conditional laws of the model through the conditional Fourier-Laplace transform.

Once the convergence of the Bermudan options has been established, we prove the convergence of American option prices in the multi-factor model (X^n, V^n) towards the prices in the limiting Volterra model, by approximating them with Bermudan option prices using classical arguments Lamberton and Pagès [1990].

We finish the chapter with some numerical implementations. We employ the multi-factor approximation (X^n, V^n) whose trajectories can be simulated. We use an implicit-explicit Euler scheme for its simulation and the classical Longstaff Schwartz algorithm [Longstaff and Schwartz, 2001] to compute the prices of American options.

Moments computation

In Chapter 3 we start with the brief study of polynomial diffusions. A polynomial diffusion is a weak solution of (26), characterized by having an affine drift and a quadratic diffusion. Polynomial diffusions are used in many financial applications for models of the term structure of variance swap rates [Filipović et al., 2016] and interest rates [Filipović et al., 2017], in credit risk [Ackerer and Filipović, 2020], for option pricing using variance reduction techniques among other applications [Cuchiero et al., 2012].

An important feature of these models is that conditional moments are given in a closed formula [Cuchiero et al., 2012, Filipović and Larsson, 2016]. Conditional moments of polynomial processes can be evaluated trough the computation of a matrix exponential. In particular, if X is a polynomial diffusion, then its moments are polynomial functions of the initial state X_0. The efficient and easy computation of their moments, leads for instance to an efficient way to price options trough density approximation techniques [Ackerer et al., 2018, Filipović et al., 2013, Heston and Rossi, 2017]. Affine diffusions are a particular case of polynomial diffusions.

Our goal is to recover some of the important properties to compute moments of polynomial diffusions in a polynomial Volterra setting. We consider as a polynomial Volterra process, a weak solution of (25) with affine drift and a quadratic diffusion. We remark that polynomial Volterra processes include the affine Volterra as a particular case.

Our main result is Theorem 3.3.3, which states that the moments of order p of an affine Volterra process are polynomials in the initial state of degree less than or equal to p. The coefficients of these polynomials are independent of the initial condition.

In this chapter, we also present some general formulas to calculate the conditional moments of affine Volterra processes in Sections 3.3.2 and 3.3.3.

In Theorem 3.4.1 we show that if we consider a sequence $X^n = (X^n_t)_{t \leq T}$ of continuous weak solutions to the stochastic Volterra equation (25) that converge weakly to X solution of (25). Under some uniformly bounded conditions over X^n, we get convergence of moments of X^n to the moments of X:

$$\mathbb{E}[P(X^n)] \to \mathbb{E}[P(X)], \quad \text{for every polynomial } P \text{ on } \mathbb{R}^d.$$

This last theorem give us a method to compute the moments of polynomial Volterra processes in an efficient way. Using the approximation technique studied in Section 1.4.2, we can construct a multi-factor approximation X^n of X. One important remark is that the factors in X^n are polynomial diffusions. Hence, moments can be computed through the evaluation of an exponential matrix. We finished the chapter with some numerical examples.

Parameter estimation

Statistical inference of diffusion processes driven by a standard Brownian motion has a long history. The problem to estimate the drift parameter if the process can be observed in continuous or in discrete time has been amply studied when the process is driven by a standard

Brownian motion (see Kutoyants [2004]Liptser and Shiryaev [2001] Basawa and Scott [1983] and the references therein).

Estimation of the drift parameter in the Ornstein-Uhlenbeck processes has been motivated by practical needs of having a better understanding of volatility modeling in a financial context. Indeed, mean-reverting process are used to model the volatility in stochastic volatility models. The Ornstein-Uhlenbeck process [Uhlenbeck and Ornstein, 1930] is a continuous Gaussian Markovian process satisfying the following stochastic differential equation:

$$X_t = \theta X_t dt + \sigma dW_t, \quad 0 \leq t \leq T, \tag{39}$$

where $\theta \in \mathbb{R}$ is the drift parameter, $\sigma \in \mathbb{R}_+$ is the volatility parameter and W a standard Brownian motion.

In this case, the variance parameter can be computed from the quadratic variation of the observed process. We can then assume that the variance parameter is known and equal to 1. There are two popular estimators for the drift parameter: the maximum likelihood and the least squares estimator. In this case both estimators coincide and are strongly consistent. The maximum likelihood estimator θ_t of θ, due to the Girsanov theorem [Liptser and Shiryaev, 2001], has the following form :

$$\hat{\theta}_t = \frac{\int_0^t X_s dX_s}{\int_0^t X_s^2 ds}. \tag{40}$$

Statistical inference when the volatility exhibits long memory has raised the interest of some practitioners. This feature is achieved replacing the Brownian motion in (39) by its fractional version. This new model is called the fractional Ornstein-Uhlenbeck process, which is the unique Gaussian process satisfying :

$$X_t = \theta \int_0^t X_s ds + \sigma dW_t^H, \quad 0 \leq t \leq T, \tag{41}$$

where W^H is the fractional Brownian motion (22) with Hurst parameter $H \in (0, 1)$. The Hurst parameter H characterizes the Hölder regularity of the trajectories. In the standard Brownian motion $H = 1/2$ the increments are independent, if $H > 1/2$ the increments are positively correlated and the trajectories are more regular than the standard Brownian motion. When

$H < 1/2$ the increments are negatively correlated and the trajectories are less regular. Estimation of the drift parameter in the fractional Ornstein-Uhlenbeck process is a challenging problem. Mainly because the fractional Brownian is neither a semimartingale nor a Markovian process for $H \neq 1/2$.

The long memory of the volatility is obtained when $H > 1/2$. In this case, there have been some works regarding the strong consistency of the maximum likelihood and the least squares estimators for continuous [Kleptsyna and Le Breton, 2002, Hu and Nualart, 2010, Belfadli et al., 2011] and discrete [Cénac and Es-Sebaiy, 2015, Es-Sebaiy and Ndiaye, 2014, Hu and Song, 2013, Xiao et al., 2011, Bishwal, 2011] observations of the trajectories of the process.

However as we mention in the introduction of this work, recently studies have demonstrated that volatility is rough. Meaning, that in reality the good choice for the Hurst parameter is $H < 1/2$. In this sense there have been some works regarding the strong consistency of the maximum likelihood and the least squares estimators for continuous (see for example Hu et al. [2019], Tudor and Viens [2007], El Machkouri et al. [2016]) and discrete (see for example Tudor and Viens [2007], Kubilius et al. [2015], Es-Sebaiy et al. [2019]) observations of the trajectories of the process.

In Chapter 4 we are interested in the statistical problem of the estimation of the drift parameter in the Volterra version of the Ornstein–Uhlenbeck process. To the best of our knowledge this is the first work addressing this problem. The Volterra Ornstein–Uhlenbeck process is a particular case of the affine Volterra processes and it is the unique strong solution of the stochastic Volterra equation:

$$X_t = \int_0^t K(t-s)(\theta X_s ds + dW_s) \quad t \geq 0, \tag{42}$$

where $\theta \in \mathbb{R}$, the constant drift is an unknown parameter. The kernel K that we consider is the fractional kernel $K(t) = t^{\alpha-1}/\Gamma(\alpha), \quad \alpha \in (1/2, 1]$, where the parameter $\alpha = H + 1/2$ governs the regularity of the trajectories of the process. Notice that when $\alpha = 1$ we recover the classical Ornstein–Uhlenbeck process.

Estimation of the drift parameter in the Volterra Ornstein-Uhlenbeck process is a challenging problem. Indeed, this process is neither a semimartingale nor Markovian as in the fractional

Ornstein-Uhlenbeck model. But, unlike in the fractional case, the kernel also appears next to the parameter we want to estimate, making the statistical inference even more cumbersome.

Our goal is to estimate the parameter θ in (42). First, we investigate its estimation when the process is observed in continuous time. The maximum likelihood estimator $\hat{\theta}_t$ of θ is given by:

$$\hat{\theta}_t = \frac{\int_0^t X_s dZ_s}{\int_0^t X_s^2 ds}, \tag{43}$$

where Z is an auxiliary process we introduce to facilitate the estimation. This process, contrary to X, is a semimartingale.

Our main result in this sense is Theorem 4.2.1, which states that the estimator in (43) is strongly consistent for any $\theta > 0$, that is

$$\lim_{t\to\infty} \hat{\theta}_t = \theta, \quad \text{a.s.}$$

Observations are often discretely sampled. Consequently, for practical purposes, statistical inference of processes observed in discrete time is of significant interest. In this sense, we study the problem of estimating the drift parameter in the Volterra Ornstein-Uhlenbeck process based on the discrete observations of the process. We follow the ideas of Kubilius et al. [2015] for the fractional Ornstein-Uhlenbeck process. We consider a discrete estimator that is similar in form to the maximum likelihood estimator. We assume that the process X is observed at points $\{t_k, n \geq 1, 0 \leq k \leq a_n\}$ and consider the following estimator:

$$\bar{\theta}_n(a) = \frac{\sum_{k=0}^{a_n-1} X_{t_k} \Delta Z_{t_k}}{\frac{1}{n}\sum_{k=0}^{a_n-1} X_{t_k}^2}, \tag{44}$$

where $\Delta Z_{t_k} = Z_{t_{k+1}} - Z_{t_k}$.

Our main result regarding discrete observations of the process is Theorem 4.3.1, which states that for $\theta \in \mathbb{R}$, the estimator $\bar{\theta}_n(a)$ in (44) is weakly consistent considering a sequence a that satisfies some limiting conditions.

Observe that the process X in (42) is a centered Gaussian process and can be written in an explicit form (see Abi Jaber et al. [2019]) as:

$$X_t = \int_0^t K_\theta(t-s)\sigma dW_s,$$

where $K_\theta(t) = t^{\alpha-1}E_{\alpha,\alpha}(\theta t^\alpha)$ and $E_{\alpha,\alpha}$ denotes the Mittag-Leffler function:

$$E_{\alpha,\alpha}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \alpha)}.$$

Therefore, in order to prove the consistency for the estimators in discrete and continuous time, a good understanding of the properties and asymptotic expansions of the Mittag-Leffler function is required.

At the end of the chapter we present numerical results based on simulation experiments for the discrete estimator $\bar{\theta}_n(a)$. We simulate the Ornstein-Uhlenbeck process based on the multi-factor approximation of Volterra processes studied in Section 1.4.2.

Chapter 1

Stochastic Volterra equations

In this chapter we are going to summarize existing results and properties of the Stochastic Volterra equations that will be useful in the following chapters. The chapter is organized as follows. We start with the definition of the stochastic Volterra equations. In Section 1.1 we give a brief summary of stochastic convolutions and resolvents, which are the key tools for the study of these stochastic equations. In Section 1.2 we present theorems to ensure the existence of solutions for these equations. Then, in Section 1.3, we introduce the affine Volterra processes, defined as solutions of stochastic Volterra equations with affine coefficients. We present the exponential-affine representations of their Fourier Laplace transform. Also we give some examples where the uniqueness of solutions for theses processes has been proved. We end this chapter characterizing the Markovian structure of stochastic Volterra equations in Section 1.4.

Notation:
$(\mathbb{R}^d)^*$ and $(\mathbb{C}^d)^*$ represent the dual spaces of $\mathbb{R}^d$ and $\mathbb{C}^d = \mathbb{R}^d + \mathrm{i}\mathbb{R}^d$, respectively. $A^\top$ denotes the transpose (ordinary, not conjugate) of a matrix A with complex entries. $\mathrm{Tr}(A)$ denotes the trace of the matrix A. $\mathbb{I}_d$ represents the $d \times d$-identity matrix. Δ_h is the shift operator: $\Delta_h f(t) = f(t+h)$ for any function f on $\mathbb{R}_+$. $|.|$ denotes the Euclidean norm on $(\mathbb{C}^d)^*$ and $\mathbb{C}^d$ and the operator norm on $\mathbb{R}^{m\times n}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a right continuous filtration $\mathbb{F} = (\mathcal{F}_t)_{t\leq 0}$. We assume that $\mathcal{F}_0$ contains the $\mathbb{P}$-null sets of $\mathcal{F}$. On this probability space, we consider a m-dimensional Brownian motion W. We call stochastic Volterra equations to the class of the

d-dimensional stochastic convolution equations of the form:

$$X_t = X_0 + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{1.1}$$

where $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ is a convolution kernel, $X_0 \in \mathbb{R}^d$ deterministic and the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ continuous.

Notice that $K \equiv \mathbb{I}_d$ corresponds to the classical stochastic differential equations. For more general kernels, stochastic Volterra equations do not fall in general in the semi-martingale and Markov setting. In order to handle this lack of Markov and semi-martingale structure, Abi Jaber et al. [2019] rely on tools from the theory of convolution Volterra integral equations (see Gripenberg et al. [1990] for a deep study of Volterra integral equations). In particular, the resolvents of the first and the second kind associated to the convolution kernel K play an important role.

1.1 Stochastic convolutions and resolvents

We introduce the convolution notation, as well as some important properties and the resolvents associated to the convolution kernel K following Abi Jaber et al. [2019]. For a measurable function K on $\mathbb{R}_+$ and a measure L on $\mathbb{R}_+$ of locally bounded variation, the convolutions $K * L$ and $L * K$ are defined for $t > 0$, by:

$$(K * L)(t) = \int_{[0,t]} K(t-s)L(ds), \qquad (L * K)(t) = \int_{[0,t]} L(ds)K(t-s), \tag{1.2}$$

whenever these expressions are well defined.

If F is a function on $\mathbb{R}_+$, we write $K * F = K * (F dt)$, that is:

$$(K * F)(t) = \int_0^t K(t-s)F(s)ds, \tag{1.3}$$

Fix $d \in \mathbb{N}$ and let M be a d-dimensional continuous local martingale. If K is $\mathbb{R}^{m\times d}$ valued for some $m \in \mathbb{N}$, the convolution:

$$(K * dM)_t = \int_0^t K(t-s)dM_s, \tag{1.4}$$

is well-defined as an Itô integral for any $t \geq 0$ such that $\int_0^t |K(t-s)|^2 d\,\mathrm{Tr}\langle M\rangle_s < \infty$. In particular, if $K \in L^2_{loc}(\mathbb{R}_+)$ and $\langle M\rangle_s = \int_0^s a_u du$ for some locally bounded process a, then (1.4) is well-defined for every $t \geq 0$. Moreover, the convolution (1.4) is associative:

$$(L * (K * dM))_t = ((L * K) * dM)_t,$$

for every $t \geq 0$, whenever $K \in L^2_{loc}\left(\mathbb{R}_+, \mathbb{R}^{m\times d}\right)$ and L a $\mathbb{R}^{n\times m}$ -valued measure on $\mathbb{R}_+$ of locally bounded variation. Taking $F \in L^1_{loc}\left(\mathbb{R}_+\right)$ we obtain:

$$(F * (K * dM))_t = ((F * K) * dM)_t.$$

In the following we are going to assume that the following condition on K holds, which ensures a continuous version of the convolution (1.4).

$$\begin{aligned} &K \in L^2_{loc}\left(\mathbb{R}_+, \mathbb{R}\right) \text{ and there is } \gamma \in (0,2] \text{ such that,} \\ &\int_0^h K(t)^2 dt = O\left(h^\gamma\right) \text{ and } \int_0^T (K(t+h) - K(t))^2 dt = O\left(h^\gamma\right) \text{ for every } T < \infty. \end{aligned} \tag{1.5}$$

Definition 1.1.1 (Resolvent of the second kind)**.** We consider a kernel $K \in L^1_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$. The resolvent, or *resolvent of the second kind* of K is defined as the unique kernel $R \in L^1_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ such that:

$$K * R = R * K = K - R.$$

Definition 1.1.2 (Resolvent of the first kind)**.** The *resolvent of the first kind* of K is defined as a $\mathbb{R}^{d\times d}$ valued measure L on $\mathbb{R}_+$ of locally bounded variation such that:

$$K * L = L * K = \mathbb{I}_d.$$

Notice that the resolvent always exists and is unique. However, the resolvent of the first kind does not always exist. In particular, if K is non-negative, non-increasing and not identically zero on $\mathbb{R}_+$, it admits a resolvent of the first kind.

Lemma 1.1.3. *Assume that $K \in L^1_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ admits a resolvent of first kind L. For any $F \in L^1_{loc}(\mathbb{R}_+, \mathbb{C}^{m\times d})$ such that $F * L$ is right continuous and of locally bounded variation one has:*

$$F = (F * L)(0)K + d(F * L) * K \quad a.e.$$

Notice that if K is continuous on $(0, \infty)$, then $\Delta_h K * L$ is right-continuous and we can use Lemma 1.1.3 with $F = \Delta_h K$, for fixed $h > 0$. Moreover if K is non-negative and L is non-increasing in the sense that $s \to L([s, s + t])$ is non-increasing for all $t \geq 0$, then $\Delta_h K * L$ in non-decreasing and is of locally bounded variation.

The following table presents some examples of kernels with their respective resolvents. We are particularly interested in the fractional kernel, as we can see in the next chapters.

	$K(t)$	$R(t)$	$L(dt)$
Constant	c	ce^{-ct}	$c^{-1}\delta_0(dt)$
Fractional	$c\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$ct^{\alpha-1}E_{\alpha,\alpha}\left(-ct^{\alpha}\right)$	$c^{-1}\frac{t^{-\alpha}}{\Gamma(1-\alpha)}dt$
Exponential	$ce^{-\lambda t}$	$ce^{-\lambda t}e^{-ct}$	$c^{-1}\left(\delta_0(dt)+\lambda dt\right)$
Gamma	$ce^{-\lambda t}\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$ce^{-\lambda t}t^{\alpha-1}E_{\alpha,\alpha}\left(-ct^{\alpha}\right)$	$c^{-1}\frac{1}{\Gamma(1-\alpha)}e^{-\lambda t}\frac{d}{dt}\left(t^{-\alpha}*e^{\lambda t}\right)(t)dt$

Table 1.1: Kernels K and their resolvents R and L of the second and first kind, respectively. δ_0 represent the Dirac measure, $E_{\alpha,\beta}(z)=\sum_{n=0}^{\infty}\frac{z^n}{\Gamma(\alpha n+\beta)}$ is the Mittag-Leffler function and the constant c may be an invertible matrix.

1.2 Existence of solutions for the stochastic Volterra equations

We introduce a subset $\mathcal{K}^2$ of $L^2_{loc}(\mathbb{R}_+,\mathbb{R})$, which will be very useful to show the existence of solutions to (1.1) restricted to $\mathbb{R}^d_+$.

Definition 1.2.1. For $K \in L^2_{loc}(\mathbb{R}_+,\mathbb{R})$, we have $K \in \mathcal{K}^2$ if it satisfies the following properties:

i) There exists $\gamma \in (0,2]$ such that

$$\int_0^h K(t)^2 dt = \mathrm{O}(h^\gamma) \quad \text{and} \quad \int_0^T (K(t+h)-K(t))^2 dt = \mathrm{O}(h^\gamma) \text{ for every } T<\infty.$$

ii) K is non-identically zero, non-negative, non-increasing, continuous on $(0,\infty)$ and admits a resolvent of first kind L. In addition, L is non-negative and

$$\text{the function } s \to L([s,s+t]) \text{ is non-increasing on } \mathbb{R}_+$$

for every $t>0$.

Remark 1.2.2. We remark that if K is completely monotone on $(0,\infty)$ (see Section 1.4.2 for completely monotone kernels) and not identically zero, then K satisfies condition ii) of Definition 1.2.1. In particular the fractional kernel where $K(t)=t^{\alpha-1}$ on $[0,T]$, with $\alpha \in (1/2,1]$ is a completely monotone kernel and also satisfies condition i) of Definition 1.2.1 with $\gamma = 2\alpha-1$.

Now, we are ready to state some existence results for the stochastic Volterra equations (1.1) (refer to Abi Jaber et al. [2019, Theorem 3.3, 3.4 and 3.6] for their proofs). A solution X is called *strong* if it is adapted to the filtration generated by the Brownian motion W. In the case of *weak* solutions X is not necessarily the case.

Theorem 1.2.3. *Assume that the components of K satisfy the condition (1.5).*

- *If b and σ are Lipschitz continuous, then (1.1) admits a unique strong continuous solution X for any initial condition $X_0 \in \mathbb{R}^d$.*
- *If K admits a resolvent of the first kind and b and σ satisfy the following linear growth condition:*

$$|b(x)| \vee |\sigma(x)| \leq c_{LG}(1+|x|), \text{ for any } x \in \mathbb{R}^d \text{ and with } c_{LG} \geq 0. \tag{1.6}$$

  Then, (1.1) admits a continuous weak solution for any initial condition $X_0 \in \mathbb{R}^d$.

In both cases, for any $p \geq 2$ and $T < \infty$, X satisfies:

$$\sup_{t\in[0,T]} \mathbb{E}\left[|X_t|^p\right] < \infty, \tag{1.7}$$

and admits Hölder continuous paths on $[0,T]$ of any order strictly less than $\gamma/2$.

Theorem 1.2.4. *Assume that:*

- *K is diagonal with scalar kernels K_i on the diagonal that belong to $\mathcal{K}$ (see Definition 1.2.1).*
- *b and σ satisfy the linear growth condition (1.6), and the boundary conditions:*

$$x_i = 0 \text{ implies } b_i(x) \geq 0 \text{ and } \sigma_i(x) = 0,$$

  where $\sigma_i(x)$ is the ith row of $\sigma(x)$.

Then, (1.1) admits an $\mathbb{R}^d_+$-valued continuous weak solution for any initial condition $X_0 \in \mathbb{R}^d_+$.

Once we have stated some general results for the existence of solutions to equations of the form (1.1), we can study the uniqueness of these solutions in a specif class of processes called affine Volterra processes.

1.3 Affine Volterra Processes

In this section we summarize the results of the Affine Volterra process presented in Abi Jaber et al. [2019]. Fix a dimension $d \geq 1$ and a state space $E \subseteq \mathbb{R}^d$. We let $b : E \to \mathbb{R}^d$ be continuous and such that $a = \sigma\sigma^\top$ is continuous and positive semidefinite. Let a and b be affine maps given by:

$$a(x) = A_0 + x_1 A_1 + \ldots x_d A_d, \tag{1.8}$$

$$b(x) = b_0 + x_1 b_1 + \ldots x_d b_d, \tag{1.9}$$

where A_i are $d-$dimensional symmetric matrices and b_i are $d-$dimensional vectors. In this context, we refer to the continuous E-valued solutions of (1.1) as affine Volterra processes.

Affine diffusions are particular cases of these processes where the convolution kernel $K = \mathbb{I}_d$. Affine processes are very tractable because their characteristic function has a closed formula, which is useful to solve pricing, calibration, hedging problems. The Fourier-Laplace transform of X_T is exponential affine in X_t, for all $t \leq T$. That is, there exist $\mathbb{C}$ and $(\mathbb{C}^d)^*$-valued functions ϕ and ψ, respectively, such that:

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \tag{1.10}$$

for a row vector $u \in i\mathbb{R}^d$, where ϕ and ψ solve the system of Riccati equations :

$$\phi(t) = \int_0^t \left(\psi(s)b_0 + \frac{1}{2}\psi(s)A_0\psi(s)^\top\right) ds, \tag{1.11}$$

$$\psi(t) = u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds. \tag{1.12}$$

with $A(u) = (uA_1u^\top, \ldots, uA_du^\top)$ and $B = (b_1 \ldots b_d)$. Since the conditional characteristic function is bounded by one, the real part of the exponent $\phi(T-t) + \psi(T-t)X_T$ has to be negative. Note that ϕ and ψ satisfy the initial conditions $\phi(0) = 0$ and $\psi(0) = u$. Using the variation of constants formula on X and ψ, one can write the Fourier-Laplace transform alternatively to (1.10) as:

$$\mathbb{E}\left[\exp\left(uX_T\right)|\mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T|\mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}\left[X_s|\mathcal{F}_t\right])\psi(T-s)^\top ds\right). \tag{1.13}$$

For general kernels, as we stated before affine Volterra processes are neither semi-martingales, nor Markov processes. Therefore in general, a formula like (1.10) doesn't hold. However in Abi Jaber et al. [2019], it is shown that (1.13) holds for affine Volterra processes. Their conditional Fourier Laplace transform is expressed as a functional of an affine Volterra process thanks to the solution of a quadratic Volterra equation called Riccati-Volterra equation as we can see in the following theorem (refer to Abi Jaber et al. [2019, Theorem 4.3] for the proof).

Theorem 1.3.1 (Conditional Fourier-Laplace functional of affine Volterra process). *Let X be an affine Volterra process and fix some $T < \infty$, $u \in (\mathbb{C}^d)^*$. Assume $\psi \in L^2\left([0,T], (\mathbb{C}^d)^*\right)$ solves the Riccati-Volterra equation*

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) K(t-s)ds. \tag{1.14}$$

Then, the process $\{Y_t, 0 \le t \le T\}$ *defined by:*

$$Y_t = Y_0 + \int_0^t \psi(T-s)\sigma(X_s)\, dW_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\,\psi(T-s)^\top ds, \tag{1.15}$$

$$Y_0 = uX_0 + \int_0^T \left(\psi(s)b(X_0) + \frac{1}{2}\psi(s)a(X_0)\,\psi(s)^\top\right) ds, \tag{1.16}$$

satisfies:

$$Y_t = \mathbb{E}\left[uX_T \mid \mathcal{F}_t\right] + \frac{1}{2}\int_t^T \psi(T-s)a\left(\mathbb{E}\left[X_s \mid \mathcal{F}_t\right]\right)\psi(T-s)^\top ds, \tag{1.17}$$

for all $0 \le t \le T$*. The process* $\{\exp(Y_t), 0 \le t \le T\}$ *is a local martingale and, if it is a martingale, one has the exponential-affine transform formula:*

$$\mathbb{E}\left[\exp(uX_T) \mid \mathcal{F}_t\right] = \exp(Y_t), \quad t \le T. \tag{1.18}$$

The Riccati-Volterra equation (1.14) is equivalent to:

$$\psi(t) = uE_B(t) + \int_0^t \frac{1}{2}A(\psi(s))\,E_B(t-s)ds. \tag{1.19}$$

where $E_B = K - R_B * K$ and R_B is the resolvent of $-KB$.

Next Lemma [Abi Jaber et al., 2019, Lemma 4.2] gives the conditional expectation of affine Volterra processes used in equation (1.17) .

Lemma 1.3.2 (Conditional expectation of affine Volterra process)**.** *Let X be an affine Volterra process. Then for all $t \le T$.*

$$\mathbb{E}[X_T|\mathcal{F}_t] = \left(\mathbb{I}_d - \int_0^T R_B(s)ds\right)X_0 + \left(\int_0^T E_B(s)ds\right)b_0 + \int_0^t E_B(T-s)\sigma(X_s)dW_s, \tag{1.20}$$

where R_B is the resolvent of $-KB$ and $E_B = K - R_B * K$. In particular,

$$\mathbb{E}[X_T] = \left(\mathbb{I}_d - \int_0^T R_B(s)ds\right)X_0 + \left(\int_0^T E_B(s)ds\right)b_0.$$

Uniqueness follows as a corollary from Theorem 1.3.1. Existence and uniqueness of global solutions for the stochastic affine equation (1.1) and its Riccati-Volterra equation (1.14) have been established for some state spaces. We can find some examples below.

Examples

The following examples can be extended in $\mathbb{R}^d$. For simplicity of the exposition we work in $\mathbb{R}$.

- **The Volterra Ornstein-Uhlenbeck process**

Let the state space be $E = \mathbb{R}$. We call the Volterra Ornstein-Uhlenbeck process the solution to the following equation:

$$X_t = X_0 + \theta \int_0^t K(t-s)(\lambda - X_s)\, ds + \int_0^t K(t-s)\sigma dW_s, \tag{1.21}$$

where θ, λ and σ are constants. In this case, we can obtain an explicit form for the process X,

$$X_t = \left(1 - \int_0^t R_\theta(s) ds\right) X_0 + \left(\int_0^t E_\theta(s) ds\right) \theta\lambda + \int_0^t E_\theta(t-s)\sigma dW_s,$$

where $E_\theta = K - R_\theta * K$ with R_θ the resolvent of θK. In particular, X_t is a Gaussian process and the Riccati-Volterra equation (1.14) has a explicit solution:

$$\psi(t) = u E_\theta(t).$$

The quadratic variation of the process Y in Theorem 1.3.1 is deterministic,

$$\langle Y \rangle_t = \int_0^t \psi^2(T-s)\sigma^2 ds.$$

Thus, the martingale condition in Theorem 1.3.1 holds and we have existence and uniqueness.

- **The Volterra square-root process**

When $\eta(x) = \eta\sqrt{x}$, we call the Volterra square-root process as the $\mathbb{R}_+$ solution of the following equation

$$X_t = X_0 + \lambda \int_0^t K(t-s)(\bar{\nu} - X_s)\, ds + \int_0^t K(t-s)\eta\sqrt{X_s} dW_s, \tag{1.22}$$

where λ, $\bar{\nu}$ and η are constants.

Lemma 1.3.3. *Assume that the following conditions hold:*

– *Inward-pointing drift condition:*

$$\lambda\bar{\nu} \geq 0.$$

– *The kernel K satisfies* (1.5) *and the shifted kernel $\Delta_h K$ satisfies condition ii) of Definition* (1.2.1) *for all $h \in [0,1]$,*

Then,

– *The stochastic Volterra equation* (1.22) *has unique* $\mathbb{R}_+$*-valued continuous weak solution for any initial condition* $X_0 \in \mathbb{R}_+$. *Moreover, the paths of* X *are Hölder continuous of any order less than* $\gamma/2$, *where* γ *corresponds to the constant in the assumption over* K (1.5).

– *For any* $u \in \mathbb{C}$ *with* $\operatorname{Re} u \leq 0$, *the associated Riccati-Volterra equation* (1.14), *which takes the form:*

$$\psi(t) = uK(t) + \int_0^t K(t-s)\left(-\lambda\psi(s) + \frac{1}{2}\eta^2\psi^2(s)\right) ds,$$

has a unique global solution $\psi \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{C})$ *with* $\operatorname{Re}\psi \leq 0$. *The martingale condition in Theorem 1.3.1 holds, as does the affine transform formula.*

- **The Volterra Heston model**

We consider an affine Volterra process $X = (\log S, V)$ with state space $\mathbb{R} \times \mathbb{R}_+$, where:

$$\frac{dS_t}{S_t} = \sqrt{V_t} dB_t \tag{1.23}$$

$$V_t = V_0 + \int_0^t K(t-s)\left(\lambda(\bar{\nu} - V_s)\, ds + \eta\sqrt{V_s} dW_s\right), \tag{1.24}$$

with parameters V_0, λ, $\bar{\nu}$ and η positive and $d\langle B, W\rangle_t = \rho dt$ with $\rho \in [-1, 1]$.

Lemma 1.3.4. *Assume that the kernel* K *satisfies* (1.5) *and the shifted kernel* $\Delta_h K$ *satisfies condition ii) of Definition* (1.2.1) *for all* $h \in [0, 1]$. *Then,*

– *The stochastic Volterra process* X *has a unique* $\mathbb{R} \times \mathbb{R}_+$*-valued continuous weak solution* $(\log S, V)$ *for any initial condition* $(\log S_0, V_0) \in \mathbb{R} \times \mathbb{R}_+$. *Moreover, the paths of* V *are Hölder continuous of any order less than* $\gamma/2$ *where* γ *corresponds to the constant in the assumption over* K (1.5).

– *For any* $u \in (\mathbb{C}^2)^*$ *such that* $\operatorname{Re} u_1 \in [0, 1]$ *and* $\operatorname{Re} u_2 \leq 0$, *the Riccati-Volterra equation* (1.14) *which becomes:*

$$\psi(t) = u_2 K(t) + \int_0^t K(t-s)\left(\frac{1}{2}\left(u_1^2 - u_1\right) + (\rho\eta u_1 - \lambda)\psi(s) + \frac{1}{2}\eta^2\psi^2(s)\right) ds,$$

has a unique global solution $\psi \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{C})$, *which satisfies* $\operatorname{Re}\psi \leq 0$. *The martingale condition in in Theorem 1.3.1 holds, as does the affine transform formula.*

– *The process* S *is a martingale.*

1.4 Markovian structure of stochastic Volterra equations

We are going to work with a more general equation that the one consider in (1.1). We fix $T > 0$ and consider the following stochastic Volterra equation:

$$X_t = u_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \in [0,T], \tag{1.25}$$

where W is a m-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $u_0 : [0,T] \to \mathbb{R}^d$ is a real-valued continuous function, $K \in L^2_{loc}([0,T], \mathbb{R}^{d\times d})$, and coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ continuous.

Notice that if $u_0(t) = X_0$, we recover equation (1.1). The function u_0 will proved useful to characterize the Markovian structure behind the stochastic Volterra equations. In this section we will study the existence of solutions to equations (1.25) as well as their stability. The stability theorem will allow us to approximate stochastic Volterra processes by Markovian semimartingales in a specific setting. Also we will study how stochastic Volterra equations can be lifted to infinite dimensional equations.

Similar existence results of continuous solutions for (1.25) as in Theorems 1.2.3 and 1.2.4 can be obtained under additional conditions over u_0 (see, Abi Jaber and El Euch [2019a, Theorems A.1 and A.2]).

Theorem 1.4.1. *Assume that K satisfies:*

There exist $\gamma > 0$ and $C > 0$ such that the following condition holds for any $t, h \geq 0$ with $t+h \leq T$,

$$|u_0(t+h) - u_0(t)|^2 + \int_0^{T-h} |K(s+h) - K(s)|^2 ds + \int_0^h |K(s)|^2 ds \leq Ch^\gamma. \tag{1.26}$$

- *If b and σ are Lipschitz continuous, then (1.25) admits a unique strong continuous solution X.*

- *If K admits a resolvent of the first kind and b and σ satisfy the linear growth condition (1.6). Then, (1.25) admits a continuous weak solution X.*

In both cases X satisfies:

$$\sup_{t\in[0,T]} \mathbb{E}\left[|X_t|^p\right] < \infty, \qquad p > 0, \tag{1.27}$$

and admits Hölder continuous paths on $[0,T]$ of any order strictly less than $\gamma/2$.

Theorem 1.4.2. *Assume that* $d = m = 1$ *and* $K \in L^2([0,T],\mathbb{R})$ *satisfies* (1.26) *and is non-negative, non-increasing, and continuous on* $(0,T]$. *We also assume that its resolvent of the first kind* L *is non-negative and non-increasing in the sense that* $0 \leq L([s,s+t]) \leq L([0,t])$ *for all* $s,t \geq 0$ *with* $s+t \leq T$.
Assume also that b *and* σ *satisfy the linear growth condition* (1.6)*, and the boundary conditions:*

$$b(0) \geq 0 \text{ and } \sigma(0) = 0,$$

Then, (1.25) *admits a non-negative continuous weak solution for any* $u_0 \in \mathcal{G}_K$*, where* $\mathcal{G}_K$ *is the set of admissible input curves defined as follows:*

$$\mathcal{G}_K = \left\{ u \in \mathcal{H}^{\gamma/2} : u(0) \geq 0 \text{ such that } \Delta_h u - (\Delta_h K * L)(0)u - d(\Delta_h K * L) * u \geq 0, \text{ for } h \geq 0 \right\}, \tag{1.28}$$

with $\mathcal{H}^\gamma = \{u : \mathbb{R}_+ \to \mathbb{R},$ *locally Hölder continuous of any order strictly smaller than* $\gamma\}$.

In the case of smooth kernels, the following proposition [Abi Jaber and El Euch, 2019b, Proposition B.3] ensures the strong existence and uniqueness of solutions to (1.25).

Proposition 1.4.3. *Assume that* $m = d = 1$ *and* u_0 *is Hölder continuous,* $K \in C^1([0,T],\mathbb{R})$ *admitting a resolvent of first kind and that there exist* $C > 0$ *and* $\eta \in [1/2, 1]$ *such that for any* $x, y \in \mathbb{R}$,

$$|b(x) - b(y)| \leq C|x-y|, \qquad |\sigma(x) - \sigma(y)| \leq C|x-y|^\eta.$$

Then, the stochastic Volterra equation (1.25) *admits a unique strong continuous solution.*

1.4.1 Stability of stochastic Volterra equations

Once we have studied the existence of solutions to (1.25) we can enunciate a stability result [Abi Jaber and El Euch, 2019b, Theorem 3.6], of the stochastic Volterra equations.

Theorem 1.4.4 (Stability of stochastic Volterra equations)**.** *We consider a* d*-dimensional stochastic Volterra equation of the form* (1.25)*, satisfying the assumptions in Theorem 1.4.1 to ensure the existence of a continuous weak solution* X. *Consider a sequence* $(X^n)_{n\geq 1}$ *of continuous weak solution to* (1.25) *with a kernel* $K^n \in L^2_{loc}([0,T],\mathbb{R}^{d\times d})$ *which admits a resolvent of the first kind,*

$$X_t^n = u_0^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s^n, \qquad t \in [0,T], \tag{1.29}$$

where W^n is a m-dimensional Brownian motion on some filtered probability space $(\Omega^n, \mathcal{F}^n, \mathbb{F}^n, \mathbb{P}^n)$, $u_0^n : [0,T] \to \mathbb{R}^d$ and K^n satisfying (1.26) for every $n \geq 1$.

Assume that there exist positive constants γ and C such that:

$$\sup_{n\geq 1}\left(|u_0^n(t+h)-u_0^n(t)|^2+\int_0^{T-h}|K^n(s+h)-K^n(s)|^2ds+\int_0^h|K^n(s)|^2ds\right)\leq Ch^{\gamma}, \quad (1.30)$$

for any $t, h \geq 0$ with $t+h \leq T$, and assume that (K^n, u_0^n) is close to (K, u_0) in the following form:

$$\int_0^T |K(s)-K^n(s)|^2 ds \to 0, \qquad u_0^n(t) \to u_0(t), \quad (1.31)$$

for any $t \in [0,T]$ as n goes to infinity. Then, the sequence $(X^n)_{n\geq 1}$ is tight for the uniform topology and any point limit X is a solution of the stochastic Volterra equation (1.25).

Theorem 1.4.4 states the convergence in law of the family $(X^n)_{n\geq 1}$ to a limiting point X solution to (1.25).

1.4.2 Completely monotone kernels

Stochastic Volterra equations can be lifted to infinite dimensional equations (see for example Cuchiero and Teichmann [2020], Abi Jaber and El Euch [2019a] and Harms and Stefanovits [2019]). This enables to characterize the Markovian structure of X in terms of infinite dimensional processes. In this section we consider kernels that have the form (1.32) where μ is a positive measure (for signed measures refer to Cuchiero and Teichmann [2020]). This idea was initiated in the works of Carmona et al. [2000] and Muravlev [2011].

Assume that the kernel $K : [0,T] \to \mathbb{R}^{d\times d}$ in the stochastic Volterra equation (1.25) is completely monotone in the sense of Gripenberg et al. [1990, Definition 5.2.3], that is K is infinitely many times differentiable, and:

$$(-1)^j K^{(j)}(t) \quad \text{is positive semi-definite for} \quad t \in [0,T] \quad \text{and} \quad j \in \mathbb{N}.$$

Bernstein's theorem states that K is completely monotone if and only if K can be expressed as the Laplace transform of a positive $d \times d$-measure μ:

$$K(t) = \int_{\mathbb{R}_+} e^{-xt}\mu(dx). \quad (1.32)$$

Remark 1.4.5. The fractional kernel $K(t)=\frac{t^{\alpha-1}}{\Gamma(\alpha)}$, with $\alpha\in(1/2,1]$, is completely monotone and its associated measure μ is given by $\mu(dx)=\frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$.

Employing this Laplace representation of the kernel K in (1.25) and applying stochastic Fubini's theorem yields the following expression for the process X:

$$X_t = u_0(t) + \int_{\mathbb{R}_+} \mu(dx) Y_t(x), \tag{1.33}$$

where

$$Y_t(x) = \int_0^t e^{-x(t-s)} \left(b(X_s)ds + \sigma(X_s)dW_s\right). \tag{1.34}$$

It is important to notice that each process $((Y_t(x))_{x\geq 0})_{t\geq 0}$ is a Markovian semimartingale, even if X is not. Its dynamics can be represented as:

$$dY_t(x) = \left(-xY_t(x) + b\left(u_0(t) + \int_{\mathbb{R}_+} \mu(dy)Y_t(y)\right)\right)dt + \sigma\left(u_0(t) + \int_{\mathbb{R}_+} \mu(dy)Y_t(y)\right)dW_t, \tag{1.35}$$

with initial condition $Y_0(x)=0$ and x ranges in the support of μ. Hence, X can be see as a mixture of mean reverting processes.

Numerical approximation

When the kernel K is completely monotone with the form (1.32) there is some literature (see Abi Jaber and El Euch [2019b], Carmona et al. [2000], Alfonsi and Kebaier [2021]), which states that the process X can have a Markovian-semimartingale approximation denoted X^n given by a weighted sum of n factors.

As before, we consider completely monotone kernels of the form:

$$K(t) = \int_{\mathbb{R}_+} e^{-xt}\mu(dx), \quad 0\leq t\leq T,$$

where μ is a positive measure. Assume now that we can find $n\in\mathbb{N}$, $(x_i)_{1\leq i\leq n}\in\mathbb{R}_+$ and some weights $(c_i)_{1\leq i\leq n}\in\mathbb{R}^{d\times d}$ such that, if we denote by δ_x the Dirac measure at point x, $\sum_{i=1}^n c_i\delta_{x_i}$ approximates the measure μ. Then, we may define an approximation of the initial kernel K by a sequence of smoothed kernels $(K^n)_{n\in\mathbb{N}}$ as follows:

$$K^n(t) = \sum_{i=1}^n c_i e^{-x_i t}, \text{ for } 0\leq t\leq T. \tag{1.36}$$

This leads to the following approximation X^n of the process X. For $n \in \mathbb{N}$ and $t \geq 0$, we have

$$X_t^n = u_0^n(t) + \sum_{i=1}^{n} c_i Y_t^{n,i}, \tag{1.37}$$

where for any $i \in \{1, ..., n\}$, Y^n is the solution of the following SDE:

$$dY_t^{n,i} \;\; = \;\; \big(- x_i Y_t^{n,i} + b(X_t^n) \big) dt + \sigma(X_t^n) dW_t^n, \qquad Y_0^{n,i} = 0 \text{ for all i} \in \{1, 2, \ldots, n\} \tag{1.38}$$

We can rewrite (1.37) as the following stochastic Volterra equation:

$$X_t^n = u_0^n(t) + \int_0^t K^n(t-s) b(X_s^n) ds + \int_0^t K^n(t-s) \sigma(X_s^n) dW_s^n, \qquad t \in [0, T]. \tag{1.39}$$

We can ask whether or not X^n is a good approximation of X. Theorem 1.4.4 gives the conditions for the convergence in law of X^n to X when (K^n, u_0^n) is close to (K, u_0).

If we have strong solutions for (1.39) as in Theorem 1.4.3, that means that all processes can be constructed in the same probability space and we can change W^n in (1.39) by W.
Remark that the processes $Y = (Y_t^{n,i})_{i \leq n}$ are Markovian semimartigales that can be simulated. Also, if the stochastic Volterra equation (1.25) is affine, then the processes $Y = (Y_t^{n,i})_{i \leq n}$ are affine in the classical sense. In the affine case, the approximation of the kernel is used to approximate the rough Heston model by Markovian semimartingales (see Abi Jaber and El Euch [2019b] and Abi Jaber [2019a]). In particular, they are able to approximate the Volterra Riccati equations associated with the characteristic function by classical Riccati equations. This leads to a faster pricing and calibration of derivatives.

In the case of the fractional kernel, Abi Jaber and El Euch [2019b, Proposition 3.3] gives a way to choose the weights c_i's and the coefficients x_i's in order to obtain the convergence of K^n to K.

1.4.3 Adjusted forward process

Another way of recovering the Markovian structure of Stochastic Volterra equations of the form (1.25) is by using the following process consider in Abi Jaber and El Euch [2019a],

$$u_t(x) \;\; = \;\; \mathbb{E}\left[X_{t+x} - \int_0^x K(x-s) b(X_{t+s}) ds \middle| \mathcal{F}_t \right], \qquad x \geq 0. \tag{1.40}$$

The process (1.40) is called the *adjusted forward process* because it has the additional integral term that is not included in the forward process. Considering that the process $\left(\int_0^r K(t-s)\sigma(X_s)dW_s\right)_{0\leq r\leq t}$ is a martingale, the process (1.40) can be expressed as:

$$u_t(x) = u_0(t+x) + \int_0^t K(t+x-s)\left(b(X_s)ds + \sigma(X_s)dW_s\right), \qquad x \geq 0, \tag{1.41}$$

this last expression let us visualize that the shifted process $X^t = (X_{t+x})_{x\geq 0}$ satisfies the same equation (1.25) replacing u_0 by u_t, that is:

$$X_x^t = u_t(x) + \int_0^x K(x-s)\left(b(X_s^t)ds + \sigma(X_s^t)dW_s^t\right), \qquad x \geq 0, \tag{1.42}$$

where $W^t := W_{t+\cdot} - W_t$. This suggests that the process X is Markovian in the infinite dimensional curve $(u_t)_{t\geq 0}$. If conditions of Theorem 1.4.2 are satisfied such that there exists a continuous weak solution to (1.25) for any admissible input curve u_0 belonging to the set $\mathcal{G}_K$ defined by (1.28). Then, since X^t in (1.42) is non-negative, one can presume that u_t also belongs to $\mathcal{G}_K$. In Abi Jaber and El Euch [2019a], it is shown that $\mathcal{G}_K$ is stochastically invariant with respect to the family $(u_t)_{t\leq 0}$.

In Abi Jaber and El Euch [2019a], the modified adjusted process is used to characterize the Markovian structure under the dynamics of the Volterra Heston model of in terms of the stock price and the adjusted forward curve $(u_t)_{t>0}$. This process will be very useful in order to price path-dependent options as Bermuda and American options.
Equation (1.41) can be seen as the mild solution of an stochastic partial differential equation. That is,

$$du_t(x) = \left(\partial_x u_t(x) + K(x)b\left(u_t(0)\right)\right)dt + K(x)\sigma\left(u_t(0)\right)dW_t, \qquad u_0 \in \mathcal{G}_K.$$

Finally, we present the relation between the two infinite dimensional objects presented here and in Section 1.4.2. Using the Laplace transform representation (1.32) for completely monotone kernels, equation (1.41) can be seen as:

$$u_t(x) = u_0(t+x) + \int_{\mathbb{R}_+} e^{-\gamma x}\mu(d\gamma)Y_t^\gamma, \qquad t, x \geq 0, \tag{1.43}$$

where Y satisfies (1.34).

Supporting results: Existence and uniqueness of Volterra equations

In this section we present some existence and uniqueness results for solutions of Volterra equations.

Theorem 1.4.6. *[Abi Jaber et al., 2019, Theorem B.1] Assume that* $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$, $g \in L^2_{loc}(\mathbb{R}_+, \mathbb{C})$, $p : \mathbb{R}_+ \times \mathbb{C} \to \mathbb{C}$ *such that* $p(.,0) \in L^1_{loc}(\mathbb{R}_+, \mathbb{C})$*. Also, assume that for all* $T \in \mathbb{R}_+$*, there exist a positive constant* Θ_T *and a function* $\Phi_T \in L^2([0,T], \mathbb{R}_+)$ *such that*

$$|p(t,x) - p(t,y)| \leq \Phi_T(t)|x-y| + \Theta_T|x-y|(|x|+|y|), \quad x, y \in \mathbb{C}, \quad t \leq T.$$

Then, the Volterra integral equation:

$$\psi = g + K * p(.,\psi), \tag{1.44}$$

has a unique non-continuable solution (ψ, T_{max})*. If* g *and* p *are real-valued, then so is* ψ.

Remark 1.4.7. A *non-continuable* solution of (1.44) is a pair (ψ, T_{max}) with $T_{max} \in (0, \infty]$ and $\psi \in L^2_{loc}([0, T_{max}), \mathbb{C})$, such that ψ satisfies (1.44) on $[0, T_{max})$ and $\|\psi\|_{L^2(0,T_{max})} = \infty$ if $T_{max} < \infty$. If $T_{max} = \infty$ we call ψ a global solution of (1.44). A non-continuable solution $(\psi, T_{\max})$ is *unique* if for any $T \in \mathbb{R}_+$ and $\widetilde{\psi} \in L^2\left([0,T], \mathbb{C}^d\right)$ satisfying (1.44) on $[0,T]$, we have $T < T_{\max}$ and $\widetilde{\psi} = \psi$ on $[0,T]$.

Now, we proceed to state some results for linear Volterra equations. We consider kernels $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ satisfying condition (1.26) with $T = \infty$. We also consider, the space $\mathcal{G}_K$:

$$\mathcal{G}_K = \{u : \mathbb{R}_+ \to \mathbb{R} : u(0) \geq 0 \text{ such that } \Delta_h u - (\Delta_h K * L)(0)u - d(\Delta_h K * L) * u \geq 0, \text{ for } h \geq 0\},$$

Theorem 1.4.8. *[Abi Jaber and El Euch, 2019b, Theorem C.1] Let* $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ *satisfying condition* (1.26) *and* $g, z, w : \mathbb{R}_+ \to \mathbb{R}$ *be continuous functions. Then, the linear Volterra equation:*

$$\chi = g + K * (z\chi + w),$$

admits a unique continuous solution χ*. Furthermore if* $g \in \mathcal{G}_K$ *and* w *is non-negative, then* χ *is non-negative and*

$$\Delta_{t0}\chi = g_{t_0} + K * (\Delta_{t_0} z \Delta_{t_0}\chi + \Delta_{t_0} w),$$

with $g_{t_0}(t) = \Delta_{t_0} g(t) + (\Delta_t K * (z\chi + w))(t_0) \in \mathcal{G}_K$*, for all* $t, t_0 \geq 0$.

Proposition 1.4.9. *[Abi Jaber and El Euch, 2019b, Corollary C.4] Let* $h_0 \in \mathbb{C}$ *and* $z, w : \mathbb{R}_+ \to \mathbb{C}$ *be continuous functions such that* $Re(z) \leq \lambda$ *for some* $\lambda \in \mathbb{R}$. *We define* $h : \mathbb{R}_+ \to \mathbb{C}$ *as the unique continuous solution of*

$$h = h_0 + K * (zh + w).$$

Then, for any $t \in [0, T]$,

$$|h(t)| \leq |h_0| + (\|w\|_{\infty,T} + \lambda |h_0|) \int_0^T E_\lambda(s) ds,$$

where $E_\lambda = K - R_\lambda * K$ *and* R_λ *is the resolvent of* $-K\lambda$.

Chapter 2

American options in the Volterra Heston model

Stochastic volatility models whose trajectories are continuous but less regular than Brownian motion, also known as rough volatility models, seem well-adapted to capture stylized features of the time series of realized volatility and of the implied volatility surface. Indeed, recent statistical studies [Bennedsen et al., 2021, Gatheral et al., 2018, Livieri et al., 2018, Fukasawa et al., 2019] demonstrate that–under multiple time scales and across many markets–the time series of realized volatility oscillates more rapidly than Brownian motion. In addition, the observed implied volatility smile for short maturities is steeper than the one obtained with classical low-dimensional diffusion models. As maturity decreases the slope at the money of the implied volatility smile obeys a power law that explodes at zero. This power law can be reproduced by rough volatility models with power kernels in the spirit of fractional Brownian motion [Alòs et al., 2007, Fukasawa, 2011, Bayer et al., 2016, Garnier and Solna, 2017, Forde and Zhang, 2017, Fukasawa, 2017, Guennoun et al., 2018, Bayer et al., 2019, Euch et al., 2019, Fukasawa, 2021]. Furthermore, these empirical discoveries are supported by micro-structural considerations because rough volatility models appear naturally as scaling limits of micro-structural pricing models with self-exciting features driven by Hawkes processes [Jaisson and Rosenbaum, 2016, El Euch et al., 2018, Dandapani et al., 2019, Tomas and Rosenbaum, 2021].

The aforementioned findings have motivated the study of various rough volatility models in the literature. Among these are the rough fractional stochastic volatility model [Gatheral et al., 2018], the rough Bergomi model [Bayer et al., 2016], and the fractional and rough Heston models [Comte et al., 2012, El Euch and Rosenbaum, 2019]. In these models, due to the absence of the semimartingale and Markov properties, even simple tasks such as pricing European options have proved challenging. Consequently, the theory of stochastic control for rough volatility models is at an early stage. Under the rough volatility paradigm, classical control problems such as linear quadratic and optimal investment problems have only been analyzed recently in Abi Jaber et al. [2021d] and Bäuerle and Desmettre [2020], Fouque and Hu [2019], Han and Wong [2020a], Abi Jaber et al. [2021c], Han and Wong [2020b], respectively. Other recent contributions to more general control settings within the rough volatility framework can be found for instance in Viens and Zhang [2019], Han and Wong [2020c], Wang et al. [2020].

In this chapter we tackle an optimal stopping problem, namely the problem of pricing American options, in the Volterra Heston model introduced in Abi Jaber and El Euch [2019a], Abi Jaber et al. [2019]. This path-dependent problem is difficult because it requires a good understanding of the conditional laws in a model where a priori the semimartingale and Markov properties do not hold. Even though we could extend parts of the analysis to more general frameworks, we concentrate on the Volterra Heston model because in this setup–as we will explain below–we can prove the necessary convergence results.

The Volterra Heston model is a generalization of the widely-known Heston model [Heston, 1993]. The dynamics of the spot variance in the Volterra Heston model are described by a stochastic Volterra equation of convolution type–specifically, the spot variance process is a Volterra square root or CIR process. When the kernel appearing in the convolution is of power-type, one obtains the now well-known rough Heston model [El Euch and Rosenbaum, 2018, 2019]. The $\mathcal{L}^2$-regularity of the kernel in the Volterra Heston model controls the Hölder regularity of the trajectories and the steepness of the implied volatility smile for short maturities. Tractability in the Volterra Heston model is a result of a semi-explicit formula for the Fourier-Laplace transform, which resembles the formula in the classical Heston model. More precisely, the Fourier-Laplace transform can be expressed in terms of the solution to a deterministic system of convolution equations of Riccati-type. This phenomenon is a particular instance of a more general law governing the structure of the Fourier-Laplace transform of what is known as Affine Volterra Processes [Abi Jaber et al., 2019, Keller-Ressel et al., 2021, Gatheral and Keller-Ressel, 2019, Cuchiero and Teichmann, 2020]. The knowledge of the Fourier-Laplace transform in the Volterra Heston model facilitates the application of Fourier-based methods in order to price European options. This circumvents the difficulties encountered in the implementation of other popular rough volatility models, such as the rough Bergomi model, where Monte-Carlo techniques [Bayer et al., 2016, Bennedsen et al., 2017, McCrickerd and Pakkanen, 2018] or Donsker-type theorems [Horvath et al., 2017] are employed to calculate prices of European options.

The numerical resolution of the Riccati convolution equations appearing in the expression of the Fourier-Laplace transform in the Volterra Heston model is, however, cumbersome due to the possibly exploding character of the associated kernel. In order to alleviate these numerical difficulties for the rough Heston model, the author in Abi Jaber [2019a] proposed a kernel-based approximation with a diffusion–high dimensional but parsimonious–model, latter named the Lifted Heston model. Despite being a semimartingale model, the Lifted Heston model is able to mimic the rough character of the trajectories and to reproduce steep volatility smiles for short maturities. The approximation of the rough Heston model with the Lifted Heston model is an example of a more general approximation technique of Volterra processes via an approximation of the kernel in Abi Jaber and El Euch [2019b] originally inspired by Carmona et al. [2000], Harms and Stefanovits [2019]. The convergence of the approximating processes and the associated prices of European-type options is guaranteed by stability results proved in Abi Jaber and El Euch [2019b] and in a more general framework in Abi Jaber et al. [2021a].

To price American options, and inspired by the approach in Abi Jaber [2019a], Abi Jaber and El Euch [2019b], we draw upon kernel-based approximations of the Volterra Heston model. In the context of the rough Heston model where the kernel is of power-type, and for the approximation scheme in Abi Jaber [2019a], the approximated models are high dimensional-diffusion models where classical simulation-based techniques, such as the Longstaff Schwartz algorithm [Longstaff and Schwartz, 2001], can be implemented. Within this framework, we can conduct an empirical study of the the convergence and behavior of Bermudan put option prices in the approximated sequence of models. The results of our numerical experiments are summarized in Section 2.4.

Our main theoretical result is Theorem 2.1.7. In the first part of the theorem, we show convergence of prices of Bermudan options in the approximating sequence of models towards the price in the original Volterra Heston model. This result is not a direct consequence of previous stability results in Abi Jaber and El Euch [2019b], Abi Jaber et al. [2021a] because of the path-dependent structure of the option. It is at this stage, and for purely theoretical reasons, that we exploit the affine structure of the model. More precisely, in order to prove the desired convergence results we first need to establish the convergence of the conditional Fourier-Laplace transforms. Once the convergence of the Bermudan option prices is established–and using classical arguments–we can prove, in the second part of Theorem 2.1.7, the convergence of American option prices by approximating them with Bermudan option prices.

It is important to mention at this point that there exist other studies of optimal stopping and American option pricing in rough or fractional models; see for instance Horvath et al. [2017], Bayer et al. [2020b], Becker et al. [2019], Goudenège et al. [2020], Bayer et al. [2020a]. To understand the novelty of our work it is crucial to point out that in general there are two levels of approximation in the resolution of an optimal stopping problem using a probabilistic approach:

(i) First, the model has to be approximated with simpler models where the trajectories can be simulated or where prices of American options can be calculated more easily. For classical diffusion models this could correspond to a classical Euler scheme for simulation or a tree-based discrete approximation. Under rough volatility, simulation is cumbersome due to the non-Markovianity of the model. There is not a unified theory about how this approximation and simulation have to be performed. For instance, in the rough Bergomi model in order to simulate the volatility process one could use hybrid schemes [Bennedsen et al., 2017]. These schemes correspond to an approximation of the power kernel by concentrating on its behavior around zero and performing a step-wise approximation away from zero. But we could also imagine schemes relying on an approximation of the fractional kernel in terms of a sum of exponentials as in Carmona et al. [2000], Harms and Stefanovits [2019]. In this work, inspired by Abi Jaber [2019a], Abi Jaber and El Euch [2019b], we use the latter approach. Regarding the approximation via discrete-type models, in Horvath et al. [2017] the authors prove a Donsker-type theorem for certain rough volatility models and apply it to perform tree-like approximations. These approximations allow them to develop tree-based algorithms, as opposed to simulation-based techniques, to price American options. The convergence of the American option prices calculated on the approximating trees towards prices in the limiting rough models, however, is not the main goal of the study.

(ii) The second approximation occurs at the level of the resolution of the optimal stopping problem for the approximated model. In the approximated model, classical techniques such as the Longstaff Schwarz algorithm, can be difficult to implement because of the high-dimensionality of the model. It is at this stage that recent studies propose novel approaches, including techniques relying on neural networks [Lapeyre and Lelong, 2019, Goudenège et al., 2020], to ease the implementation. It is also important to mention at this point the study in Bayer et al. [2020a], where the authors propose an approximation of American option prices using penalized versions of the BSPDE satisfied by the value function of the problem. A deep learning-based method is used to approximate the solutions of these penalized BSPDEs.

The present work does not focus on the second level of the approximation. For this part, in our numerical experiments we employ classical simulation-based techniques and in particular the Longstaff Schwarz algorithm over a low dimensional space of functions. Our study mainly focuses on the first level of the approximation. More precisely, we concentrate on the convergence of the prices in the approximating model towards the prices in the limiting Volterra model. This point has not been addressed in the previous literature and is what distinguishes our work from others on American options under the rough volatility paradigm. To prove this convergence in our framework and with our kernel-based approximation approach, we appeal to the particular affine structure of the Volterra Heston model, which explains our choice of setting. One could extend some of the results to other settings as long as the results regarding the convergence of the conditional Fourier-Laplace transform remain valid. Beyond the affine paradigm, for instance for the rough Bergomi model, this question falls outside the scope of our work and it is an interesting topic for future research.

This work is based on Chevalier et al. [2021]. The rest of this chapter is organized as follows. In Section 2.1 we introduce the setup and we state our main result of convergence. Section 2.2 contains the results on the adjusted forward process and the conditional Fourier-Laplace transform necessary for the proof of the main theorem. The proof of the main theorem, namely Theorem 2.1.7, is presented in Section 2.3. In Section 2.4, within the framework of the rough

Heston model, we give a numerical illustration of the convergence and behavior of Bermudan put option prices. Section 2.5.1 explains some properties of the Riccati equations appearing in the expression of the conditional Fourier-Laplace transform. In Section 2.5.2 we provide results on the kernel approximation which guarantee certain hypotheses appearing in our main theorem.

Notation:

$\mathcal{H}^\beta = \mathcal{H}^\beta(\mathbb{R}_+, \mathbb{R})$ represents the space of real-valued functions on $\mathbb{R}_+$ that are Hölder continuous of any order strictly less than β. We denote by $\mathcal{C}(X, Y)$ the space of real-valued continuous functions from X to Y, $\mathcal{C}(X)$ if $Y = \mathbb{R}$ and $\mathcal{C}$ if in addition $X = \mathbb{R}$.

$\mathcal{C}_c$ denotes the space of functions with compact support, $\mathcal{C}_b$ the space of continuous and bounded functions and $\mathcal{C}_b^2$ the space of functions in $\mathcal{C}^2$ whose derivatives are bounded. $\mathcal{B}$ (resp. $\mathcal{B}_c$) denotes the space of bounded functions (bounded functions with compact support). $\Delta_\varepsilon f(\cdot) = f(\cdot + \varepsilon)$ is the shift-operator. We denote $\mathbf{1}$ the Indicator function and $\mathcal{L}^2_{loc}$ the space $\mathcal{L}^2_{loc}(\mathbb{R}_+, \mathbb{R})$.

2.1 Setup and main result

2.1.1 The model

We consider a Volterra Heston stochastic volatility model as in Abi Jaber and El Euch [2019a], Abi Jaber et al. [2019]. In this model, under a risk-neutral measure, the asset's log prices X and spot variance process V are

$$\begin{aligned} X_t &= X_0 + \int_0^t \left(r - \frac{V_s}{2} \right) \mathrm{d}s + \int_0^t \sqrt{V_t} \left(\rho \, \mathrm{d}W_s + \sqrt{1-\rho^2} \, \mathrm{d}W_s^\perp \right), \\ V_t &= v_0(t) - \lambda \int_0^t K(t-s) V_s \, \mathrm{d}s + \eta \int_0^t K(t-s) \sqrt{V_s} \, \mathrm{d}W_s. \end{aligned} \tag{2.1}$$

In these equations, $X_0 \in \mathbb{R}$ is the initial log price, $(W, W^\perp)$ is a two-dimensional Brownian motion, r is the risk-free rate, and $\rho \in [-1, 1]$ is a correlation parameter. The variance process V is a Volterra square root process. The constant $\lambda \geq 0$ is a parameter of mean reversion speed and $\eta \geq 0$ is the volatility of volatility. The kernel K is in $\mathcal{L}^2_{loc}$ and the function v_0 is in $\mathcal{C}(\mathbb{R}_+)$. Observe that–for fixed X_0, interest rate r, and correlation parameter ρ–the log prices X are completely determined by the variance process V and the Brownian motion $(W, W^\perp)$. Proposition 2.1.3 gives sufficient conditions ensuring the existence and uniqueness of weak solutions to the stochastic Volterra equation of the variance process.

Following the setting in Abi Jaber et al. [2019], we introduce a subset $\mathcal{K}$ of $\mathcal{L}^2_{loc}$ in which we will consider the kernels K. We shall fix a constant $\gamma \in (0, 2]$ and a locally bounded function $c_K : \mathbb{R}_+ \to \mathbb{R}_+$.

Definition 2.1.1. Let $K \in \mathcal{L}^2_{loc}$. We write $K \in \mathcal{K}$ if the following holds:

1. $$\int_0^{\varepsilon} |K(t)|^2 \, \mathrm{d}t + \int_0^{T-\varepsilon} |K(t+\varepsilon) - K(t)|^2 \, \mathrm{d}t \leq c_K(T)\varepsilon^{\gamma} \tag{2.2}$$

   for every $T > 0$ and $0 < \varepsilon \leq T$.

2. K is non-identically zero, non-negative, non-increasing, continuous on $(0, \infty)$ and admits a so-called resolvent of first kind L.[1] In addition, L is non-negative and

   $$\text{the function } s \to L([s, s+t]) \text{ is non-increasing on } \mathbb{R}_+$$

   for every $t > 0$.

We specify the space $\mathcal{G}_K$, introduced in (1.28), of functions in which we will take the functions v_0. For a given kernel $K \in \mathcal{K}$, with associated constant γ as in (2.2), let

$$\mathcal{G}_K = \{g \in \mathcal{H}^{\gamma/2} : \; g(0) \geq 0, \; \Delta_\varepsilon g - (\Delta_\varepsilon K * L)(0) g - \mathrm{d}(\Delta_\varepsilon K * L) * g \geq 0 \text{ for all } \varepsilon \geq 0\}. \tag{2.3}$$

Throughout our study we will make the following assumption.

Assumption 2.1.2. *The kernel K and the function v_0 satisfy:*

1. *$K \in \mathcal{K}$ and $\Delta_\varepsilon K$ satisfies 2 in Definition 2.1.1 for all $\varepsilon \geq 0$.*

2. *$v_0 \in \mathcal{G}_K$.*

The existence and uniqueness in law for the stochastic Volterra equation of the variance process in (2.1) is guaranteed by the following proposition.

[1] This is a real-valued measure L of locally bounded variation on $\mathbb{R}_+$ such that $K * L = 1$. Refer to Section 1.1 for a study of stochastic convolutions and resolvents.

Proposition 2.1.3. *Suppose that Assumption 2.1.2 holds. Then the stochastic Volterra equation for the variance process V in (2.1) has a unique $\mathbb{R}_+$–valued weak solution. Furthermore, the trajectories of V belong to $\mathcal{H}^{\gamma/2}$ and given $p \geq 1$*

$$\sup_{t\in[0,T]} \mathbb{E}[|V_t|^p] \leq c, \quad T > 0, \tag{2.4}$$

where $c < \infty$ is a constant that only depends on $p, T, \lambda, \eta, \gamma, c_K$ and $\|v_0\|_{\mathcal{C}[0,T]}$.

Proof. This result follows from Abi Jaber and El Euch [2019a, Theorems 2.1 and 2.3], with the exception of the last assertion on the bound (2.4). Following the argument in the proof of Abi Jaber et al. [2019, Lemma 3.1], this bound can be shown to depend on p, T, λ, η, $\|v_0\|_{\mathcal{C}[0,T]}$ and $\mathcal{L}^2$–continuously on $K|_{[0,T]}$. Note that, thanks to the Fréchet-Kolmogorov theorem, the set of restrictions $K|_{[0,T]}$ of non-increasing kernels satisfying the property (2.2) for a given c_K and γ is relatively compact in $\mathcal{L}^2(0,T)$. Maximizing the bounds over all such K yields a bound $c < \infty$ that only depends on $p, T, \lambda, \eta, \gamma, c_K$ and $\|v_0\|_{\mathcal{C}[0,T]}$. □

The theoretical results of this study are stated for general kernels K and functions v_0 satisfying Assumption 2.1.2. This is convenient in order to keep the notation simple. It is also in tune with forward-type stochastic volatility models, such as the rough Bergomi model [Bayer et al., 2016]. Indeed, thanks to (2.4), taking expectations in the equation for the variance process in (2.1) yields the following relation between the function v_0 and the initial forward-variance curve $(\mathbb{E}[V_t])$

$$v_0(t) = \mathbb{E}[V_t] + \lambda \int_0^t K(t-s)\mathbb{E}[V_s]\,\mathrm{d}s.$$

For the numerical illustrations in Section 2.4 we will use the setting of the rough Heston model [El Euch and Rosenbaum, 2019], which we summarize in the following example.

Example 2.1.4. In the rough Heston model, the kernel K is a fractional kernel

$$K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)} \tag{2.5}$$

with $\alpha = H + \frac{1}{2}$, $H \in \left(0, \frac{1}{2}\right]$, and the function v_0 is of the form

$$v_0(t) = V_0 + \lambda\overline{\nu} \int_0^t K(s)\,\mathrm{d}s, \tag{2.6}$$

where $V_0 \geq 0$ is an initial variance and $\overline{\nu} \geq 0$ is a long term mean reversion level. Assumption (2.1.2), with $\gamma = 2\alpha - 1 = 2H$, holds in this framework thanks to Abi Jaber et al. [2019, Examples 2.3 and 6.2] and Abi Jaber and El Euch [2019a, Example 2.2].

Assume that Assumption 2.1.2 holds. Let $\mathbb{P}$ be the probability measure and $\mathbb{F} = (\mathcal{F}_t)$ be the filtration of the stochastic basis associated to the weak solution (X, V) to (2.1). Suppose that $f \in \mathcal{C}_b(\mathbb{R})$. Our goal is to determine the value process $(P_t)_{0\le t\le T}$ of the American option with payoff process $(f(X_t))_{0\le t\le T}$. We know that P is given by

$$P_t = \operatorname{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)} f(X_\tau)|\mathcal{F}_t\right], \quad 0 \le t \le T, \tag{2.7}$$

where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$ and $\mathcal{T}_{t,T}$ denotes the set of $\mathbb{F}$-stopping times taking values in $[t, T]$. In order to calculate American option prices, the financial model has to be approximated by more tractable models. In this work, we will consider approximations of the Volterra Heston model resulting from $\mathcal{L}^2$-approximations of the kernel as presented in Section 1.4.2. In the next section, we describe the approximation procedure.

2.1.2 Approximation of the kernel and the Volterra Heston model

We consider a sequence of kernels $(K^n)_{n\ge1}$ in $\mathcal{L}^2_{loc}$ and functions $(v_0^n)_{n\ge1}$ in $\mathcal{C}(\mathbb{R}_+)$. We make the following assumption.

Assumption 2.1.5. *The kernels $(K^n)_{n\ge1}$ and the functions $(v_0^n)_{n\ge1}$ satisfy:*

1. *There exist a constant $\gamma \in (0, 2]$ and a locally bounded function $c_K : \mathbb{R}_+ \to \mathbb{R}_+$ such that K^n satisfies 1 for all $n \ge 1$ in Definition 2.1.1.*
2. *$\Delta_\varepsilon K^n$ satisfies 2 in Definition 2.1.1 for all $\varepsilon \ge 0$ and $n \ge 1$.*
3. *K^n converges to K in $\mathcal{L}^2_{loc}$.*
4. *$v_0^n \in \mathcal{G}_{K^n}$, with the constant γ of 1, for all $n \ge 1$ and v_0^n converges to v_0 in $\mathcal{C}(\mathbb{R}_+)$.*

According to Proposition 2.1.3, under Assumption 2.1.5, for each $n \ge 1$ there exists a unique weak solution (X^n, V^n) to

$$\begin{aligned} X_t^n &= X_0 + \int_0^t \left(r - \frac{V_s^n}{2}\right) \mathrm{d}s + \int_0^t \sqrt{V_t^n}\left(\rho\, \mathrm{d}W_s^n + \sqrt{1-\rho^2}\, \mathrm{d}W_s^{n,\perp}\right), \\ V_t^n &= v_0^n(t) - \lambda \int_0^t K^n(t-s) V_s^n\, \mathrm{d}s + \eta \int_0^t K^n(t-s)\sqrt{V_s^n}\, \mathrm{d}W_s^n. \end{aligned} \tag{2.8}$$

Furthermore, given $p \ge 1$

$$\sup_{n\ge1} \sup_{t\in[0,T]} \mathbb{E}^n[|V_t^n|^p] \le c, \quad T > 0, \tag{2.9}$$

with a constant $c < \infty$ which can be chosen to depend only on $p, T, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1} \|v_0^n\|_{\mathcal{C}[0,T]}$, and where $\mathbb{E}^n$ denotes the expectation in the respective probability space. Moreover, the argument in the proof of Abi Jaber and El Euch [2019b, Theorem 3.6] shows that

$$(X^n, V^n) \text{ converges in law to } (X, Y) \text{ in } \mathcal{C}(\mathbb{R}_+, \mathbb{R}^2), \quad \text{as } n \to \infty. \tag{2.10}$$

This is a consequence of a more general result proven in Proposition 2.2.3.

For completely monotone kernels[2], an approximation with a sum of exponentials is natural. We briefly explain this procedure below.

Approximation with a sum of exponentials

Assume that the kernel K is completely monotone. By Bernstein's theorem this is equivalent to the existence of a non-negative Borel measure μ on $\mathbb{R}_+$ such that

$$K(t) = \int_{\mathbb{R}_+} \mathrm{e}^{-xt} \mu(\mathrm{d}x). \tag{2.11}$$

As in Abi Jaber and El Euch [2019b] and Carmona et al. [2000], Harms and Stefanovits [2019], an approximation of the measure μ in (2.11) with a weighted sum of Dirac measures

$$\mu^n = \sum_{i=1}^n c_i^n \delta_{x_i^n} \tag{2.12}$$

yields a candidate approximation of the kernel

$$K^n(t) = \int_{\mathbb{R}_+} \mathrm{e}^{-xt} \mu^n(\mathrm{d}x) = \sum_{i=1}^n c_i^n \mathrm{e}^{-x_i^n t}. \tag{2.13}$$

The kernels $(K^n)_{n\geq 0}$ are completely monotone. If in addition they are not identically zero, as explained in Abi Jaber et al. [2019, Example 6.2], condition 2 in Assumption 2.1.5 holds.

The representation (2.13) yields the following factor-representation for the Volterra equation (2.8) satisfied by the variance process V^n

$$\begin{aligned} V_t^n &= v_0^n(t) + \sum_{i=1}^n c_i^n Y_t^{n,i}, \\ Y_t^{n,i} &= \int_0^t (-x_i^n Y_s^{n,i} - \lambda V_s^n) \,\mathrm{d}s + \int_0^t \eta \sqrt{V_s^n} \,\mathrm{d}W_s^n, \quad i = 1, \dots, n. \end{aligned} \tag{2.14}$$

This representation is convenient because the process $(Y^{n,i})_{i=1}^n$ is an n-dimensional Markov process with an affine structure. This observation, together with the convergence in (2.10), was

[2] K is completely monotone if $(-1)^m \frac{\mathrm{d}^m}{\mathrm{d}t^m} K(t) \geq 0$ for all non-negative integers m.

exploited in Abi Jaber and El Euch [2019b] in order to approximate European option prices in the rough Heston models employing Fourier methods. The affine structure will also play a crucial role in our study.

We now describe a natural way to determine the weights c_i^n and the points x_i^n. Let $(\eta_i^n)_{i=0}^n$ be a strictly increasing sequence in $[0,\infty)$ and define c_i^n and x_i^n as the mass and the center of mass of the interval $[\eta_{i-1}^n, \eta_i^n)$, i.e.

$$\begin{aligned} c_i^n &= \int_{[\eta_{i-1}^n, \eta_i^n)} \mu(\mathrm{d}x) = \mu([\eta_{i-1}^n, \eta_i^n)) \\ c_i^n x_i^n &= \int_{[\eta_{i-1}^n, \eta_i^n)} x\mu(\mathrm{d}x), \quad i = 1, \ldots, n. \end{aligned} \tag{2.15}$$

In Section 2.5.2 we provide sufficient conditions on the measure μ and the partitions $(\eta_i^n)_{i=0}^n$ that imply condition 1 in Assumption 2.1.5.

For the numerical illustrations in Section 2.4 we will use a fractional kernel and a geometric partition which we present in the following example.

Example 2.1.6. The fractional kernel (2.5) is completely monotone and in this case

$$\mu(\mathrm{d}x) = \frac{x^{-\alpha}}{\Gamma(1-\alpha)\Gamma(\alpha)}\,\mathrm{d}x.$$

Following Abi Jaber [2019a], we consider the geometric partition $(\eta_i^n)_{i=0}^n$ given by $\eta_i^n = r_n^{i-\frac{n}{2}}$, for $r_n > 1$ such that

$$r_n \downarrow 1 \quad \text{and} \quad n\log r_n \to \infty, \quad \text{as } n \to \infty.$$

In this setting, the vectors (c_i^n) and (x_i^n) in (2.15) take the form

$$c_i^n = \frac{(r_n^{1-\alpha}-1)}{\Gamma(\alpha)\Gamma(2-\alpha)} r_n^{(1-\alpha)(i-1-n/2)}, \qquad x_i^n = \frac{1-\alpha}{2-\alpha}\frac{r_n^{2-\alpha}-1}{r_n^{1-\alpha}-1} r_n^{i-1-n/2}, \qquad i = 1, \ldots, n. \tag{2.16}$$

Like in Example 2.1.4, along with the kernels $(K^n)_{n\geq 1}$, we consider functions $(v_0^n)_{n\geq 1}$ given by

$$v_0^n(t) = V_0 + \lambda\overline{\nu}\int_0^t K^n(s)\,\mathrm{d}s.$$

Under this framework Assumption 2.1.5 holds[3]. Indeed, Remark 2.5.4 in Section 2.5.2 shows that condition 1 holds. As explained in Abi Jaber et al. [2019, Example 6.2], condition 2 is a consequence of the complete monotonicity of K^n, $n \geq 1$. Condition 3 is shown in Abi Jaber

[3] In the case of a uniform partition $\eta_i^n = i\pi_n$, conditions that ensure 1-3 in Assumption 2.1.5 are studied in Abi Jaber and El Euch [2019b].

[2019a, Lemma A.3]. This convergence and the considerations in Example 2.1.4 imply condition 4.

With the setup of Example 2.1.6, since K^n belongs to $\mathcal{C}[0,\infty)$ and Assumption 2.1.5 holds, Abi Jaber and El Euch [2019b, Proposition B.3] implies that, for each $n \geq 1$, there exists a unique *strong solution* (X^n, V^n) to (2.8). Since in addition the factor process $(Y^{n,i})_{i=1}^n$ in (2.14) is a diffusion, classical discretization schemes can be used in order to simulate the trajectories of the variance and log prices. Relying on this observation, the numerical study in Section 2.4 uses a simulation-based method in order to approximate American option prices in the rough Heston model. The convergence of the method is a consequence of the main theoretical findings of our study, which we present in the next section.

2.1.3 Main result

We start by approximating the American option value process P in (2.7) with Bermudan option prices. More precisely, given a positive integer N, $T \geq 0$, a partition $(t_i)_{i=0}^N$ of $[0,T]$ with mesh π_N, and $t \in [0,T]$, we denote by $\mathcal{T}_{t,T}^N$ the set of $\mathbb{F}$-stopping times taking values in $[t,T] \cap \{t_0, ..., t_N\}$. For any $N \geq 1$, the Bermudan value process is then defined by

$$P_t^N = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^N} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)} f(X_\tau) | \mathcal{F}_t\right], \quad 0 \leq t \leq T. \tag{2.17}$$

In addition, given $(X^n, V^n)_{n\geq 1}$ weak solutions to (2.8), we define the corresponding American option prices

$$P_t^n = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^n} \mathbb{E}^n \left[\mathrm{e}^{-r(\tau-t)} f(X_\tau^n) | \mathcal{F}_t^n\right], \quad 0 \leq t \leq T \tag{2.18}$$

and Bermudan option prices

$$P_t^{N,n} = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^{N,n}} \mathbb{E}^n \left[\mathrm{e}^{-r(\tau-t)} f(X_\tau^n) | \mathcal{F}_t^n\right], \qquad 0 \leq t \leq T. \tag{2.19}$$

In the previous definitions, $(\mathcal{F}_t^n)$ is the filtration and $\mathbb{E}^n$ is the expectation on the stochastic basis associated to the weak solution to (2.8). The sets $\mathcal{T}_{t,T}^n$, $\mathcal{T}_{t,T}^{N,n}$ are defined similarly to $\mathcal{T}_{t,T}$, $\mathcal{T}_{t,T}^N$ on this stochastic basis.

Theorem 2.1.7 below is our main theoretical result. It implies, in particular, that the approximated American option prices P_0^n converge to the prices P_0 in the original Volterra Heston model.

Theorem 2.1.7. *Suppose that Assumptions 2.1.2 and 2.1.5 hold. Let (X, V) and (X^n, V^n) be the unique weak solutions to (2.1) and (2.8), respectively. For a function $f \in \mathcal{C}_b(\mathbb{R})$ define P, P^N, P^n and $P^{N,n}$ as in (2.7), (2.17), (2.18) and (2.19), respectively. Then*

$$P_{t_i}^{N,n} \text{ converges in law to } P_{t_i}^{N} \text{ as } n \to \infty, \quad N \geq 1,\, 0 \leq i \leq N. \tag{2.20}$$

Moreover, if $f \in \mathcal{C}_b^2(\mathbb{R})$ we have

$$\lim_{\pi_N \to 0} \sup_{n \geq 1} |P_0^{N,n} - P_0^n| = \lim_{\pi_N \to 0} |P_0^N - P_0| = 0 \tag{2.21}$$

and as a result

$$\lim_{n \to \infty} P_0^n = P_0. \tag{2.22}$$

Remark 2.1.8. For American put option prices, the convergence stated in (2.22) can be deduced by approximating the payoff function in $\mathcal{C}_b(\mathbb{R})$ with functions $(f_n)_{n\geq 1}$ in the space $\mathcal{C}_b^2(\mathbb{R})$.

The proof of Theorem 2.1.7 is based on the study of the adjusted forward variance processes and the associated Fourier-Laplace transforms, which constitute the main topic of the next section.

2.2 Conditional Fourier-Laplace transform

2.2.1 Adjusted forward process

In this section we study the adjusted forward process introduced in Section 1.4.3. This infinite-dimensional process was studied in Abi Jaber and El Euch [2019a] to characterize the Markovian structure of the Volterra Heston model (2.1). The adjusted forward process is very useful in order to study path-dependent options such as Bermudan and American options because, as we will see in Section 2.2.2, it allows us to better understand the conditional laws of the underlying process by means of the conditional Fourier-Laplace transform.

Assume that Assumption 2.1.2 holds. Let $\mathbb{P}$ be the probability measure and $\mathbb{F} = (\mathcal{F}_t)$ be the filtration of the stochastic basis associated to the weak solution (X, V) to (2.1). The adjusted forward process (v_t) of V is

$$v_t(\xi) = \mathbb{E}\left[V_{t+\xi} + \lambda \int_0^{\xi} K(\xi - s) V_{t+s}\, \mathrm{d}s \,\middle|\, \mathcal{F}_t \right], \quad \xi \geq 0.^{4} \tag{2.23}$$

[4] We called (v_t) the *adjusted* forward process to distinguish it from the classical Musiela parametrization of the forward process $(\mathbb{E}[V_{t+\cdot}|\mathcal{F}_t])$.

In particular, the variance process is embedded in the adjusted forward process because $v_t(0) = V_t$. Notice that, thanks to (2.4), the process $\left(\int_0^r K(t-s)\sqrt{V_s}\,\mathrm{d}W_s\right)_{0\leq r\leq t}$ is a martingale, and we can rewrite the adjusted forward process as

$$v_t(\xi) = v_0(t+\xi) + \int_0^t K(t+\xi-s)\left[-\lambda V_s\,\mathrm{d}s + \eta\sqrt{V_s}\,\mathrm{d}W_s\right], \quad \xi \geq 0. \tag{2.24}$$

Moreover, as shown in Abi Jaber and El Euch [2019a, Theorem 3.1], $v_t \in \mathcal{G}_K$ for all $t \geq 0$, i.e. $\mathcal{G}_K$ is stochastically invariant with respect to (v_t).

Similarly, if Assumption 2.1.5 holds, we can define the adjusted forward process for the approximating sequence $(V^n)_{n\geq 1}$ by

$$\begin{aligned} v_t^n(\xi) &= \mathbb{E}^n\left[V_{t+\xi}^n + \lambda\int_0^\xi K^n(\xi-s)V_{t+s}^n\,\mathrm{d}s\middle|\mathcal{F}_t^n\right] \\ &= v_0^n(t+\xi) + \int_0^t K^n(t+\xi-s)\left[-\lambda V_s^n\,\mathrm{d}s + \eta\sqrt{V_s^n}\,\mathrm{d}W_s^n\right], \quad \xi \geq 0, \end{aligned} \tag{2.25}$$

and we have $v_t^n(0) = V_t^n$ and $v_t^n \in \mathcal{G}_{K^n}$, for all $t \geq 0$ and $n \geq 1$.

We start with a lemma regarding the regularity for the approximated adjusted forward processes v^n, $n \geq 1$.

Lemma 2.2.1. *Let $T, M \geq 0$ and $p > \max\{2, 4/\gamma\}$. Suppose that Assumption 2.1.5 holds and for $n \geq 1$ define the processes*

$$\tilde{v}_t^n(\xi) = v_t^n(\xi) - v_0^n(t+\xi)$$

with v^n as in (2.25). Then

$$\mathbb{E}^n[|\tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi)|^p] \leq C(\max(|t-s|, |\xi-\xi'|))^{p\gamma/2}, \quad (s,\xi),(t,\xi') \in [0,T]\times[0,M],$$

where C is a constant that only depends on $p, T, M, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1}\|v_0^n\|_{\mathcal{C}[0,T]}$. As a consequence $(\tilde{v}_t^n(\xi))_{(t,\xi)\in[0,T]\times[0,M]}$ admits an α-Hölder continuous version for any $\alpha < \frac{\gamma}{2}$. Moreover, for this version and for $\alpha < \frac{\gamma}{2} - \frac{2}{p}$ we have

$$\mathbb{E}^n\left[\left(\sup_{(t,\xi')\neq(s,\xi)\in[0,T]\times[0,M]}\frac{|\tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi)|}{|(t-s,\xi'-\xi)|^\alpha}\right)^p\right] < c, \tag{2.26}$$

where $c < \infty$ is a constant that only depends on $p, \alpha, T, M, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1}\|v_0^n\|_{\mathcal{C}[0,T]}$.

Proof. Thanks to (2.25), we have for $s \leq t$ and $\xi, \xi' \leq M$

$$\begin{aligned} \tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi) &= \tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi') + \tilde{v}_s^n(\xi') - \tilde{v}_s^n(\xi) \\ &= \int_0^s (K^n(t+\xi'-u) - K^n(s+\xi'-u))\,\mathrm{d}Z_u^n + \int_s^t K^n(t+\xi'-u)\,\mathrm{d}Z_u^n \\ &\quad + \int_0^s (K^n(s+\xi'-u) - K^n(s+\xi-u))\,\mathrm{d}Z_u^n \end{aligned}$$

where $Z_t^n = -\lambda \int_0^t V_s^n \,\mathrm{d}s + \eta \int_0^t \sqrt{V_s^n}\,\mathrm{d}W_s^n$. From this point onwards, using Assumption 2.1.5 and the bound (2.9), the argument is analogous to the proof of Abi Jaber et al. [2019, Lemma 2.4] and it is based on successive applications of Jensen and Burkhölder-Davis-Gundy inequalities, and Kolmogorov's continuity theorem; see Revuz and Yor [1999, Theorem I.2.1]. □

Remark 2.2.2. As an immediate consequence of Lemma 2.2.1, if Assumption 2.1.5 holds then

$$\sup_{n\geq 1} \mathbb{E}^n \Big[\sup_{t\in[0,T]} V_t^n \Big] \leq c, \tag{2.27}$$

where $c < \infty$ is a constant that only depends on $p, T, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1} \|v_0^n\|_{\mathcal{C}[0,T]}$.

We are now able to establish the convergence of the approximated adjusted forward process in the next proposition.

Proposition 2.2.3. *Suppose that Assumptions 2.1.2 and 2.1.5 hold.Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (2.23) (resp. (2.25)). Then, as n goes to infinity, $(X_t^n, v_t^n(\xi))_{(t,\xi)\in\mathbb{R}_+^2}$ converges in law to $(X_t, v_t(\xi))_{(t,\xi)\in\mathbb{R}_+^2}$ in $\mathcal{C}(\mathbb{R}_+^2, \mathbb{R}^2)$.*

Proof. This proof is similar to the proof of Abi Jaber and El Euch [2019b, Theorem 3.6 and Proposition 4.2]. We include a short explanation for completeness. Lemma 2.2.1 and Assumption 2.1.54 imply tightness for the uniform topology of the triple (X^n, v^n, Z^n), where $Z_t^n = -\lambda \int_0^t V_s^n \,\mathrm{d}s + \eta \int_0^t \sqrt{V_s^n}\,\mathrm{d}W_s^n$. Suppose that (X, v, Z) is a limit point. Thanks to (2.25) and Abi Jaber et al. [2021a, Lemma 3.2], we have

$$\begin{aligned}
1 * v^n(\xi) &= 1 * v_0^n(\xi + \cdot) + 1 * (\Delta_\xi K^n * dZ^n) \\
&= 1 * v_0^n(\xi + \cdot) + \Delta_\xi K^n * Z^n \\
&= 1 * v_0^n(\xi + \cdot) + \Delta_\xi K * Z^n + (\Delta_\xi K - \Delta_\xi K^n) * Z^n, \quad \xi \geq 0.
\end{aligned} \tag{2.28}$$

Assumption 2.1.5 and the convergence in law of (v^n, Z) towards (v, Z) yield

$$1 * v(\xi) = 1 * v_0(\xi + \cdot) + \Delta_\xi K * Z \quad \xi \geq 0.$$

One can show, as in Abi Jaber and El Euch [2019b, Theorem 3.6], that Z is of the form $Z_t = -\lambda \int_0^t V_s \,\mathrm{d}s + \eta \int_0^t \sqrt{V_s}\,\mathrm{d}W_s$ for some Brownian motion W, where $V = v(0)$. Once again, Abi Jaber et al. [2021a, Lemma 3.2] implies that

$$v_t(\xi) = v_0(\xi + t) + (\Delta_\xi K * dZ)_t, \quad t, \xi \geq 0.$$

Hence, $V = v(0)$ is the (unique) weak solution to the stochastic Volterra equation in (2.8) and v is the associated adjusted forward process. Furthermore, one can prove that (X, V) is the (unique) weak solution of (2.1). □

2.2.2 Conditional Fourier-Laplace transforms

This section studies the conditional Fourier-Laplace transform of the log prices and the adjusted forward variance process in the Volterra Heston model based on previous considerations in Keller-Ressel et al. [2021], Abi Jaber and El Euch [2019a], Cuchiero and Teichmann [2020]. The results of this section will be useful to establish the convergence of Bermudan option prices in the approximated models to the Bermudan option prices in the original model, i.e. (2.20) in Theorem 2.1.7, using a dynamic programming approach.

We start by introducing some notation. For a kernel $K \in \mathcal{K}$ define

$$\mathcal{G}_K^* = \left\{ h \in \mathcal{B}_c(\mathbb{R}_+, \mathbb{C}) : t \mapsto -\mathrm{Re}\left(\int_0^\infty h(\xi) K(t+\xi)\,\mathrm{d}\xi \right) \in \mathcal{G}_K \right\} \tag{2.29}$$

with $\mathcal{G}_K$ as in (2.3). This space is the *dual space* that we will consider in the computation of the Fourier-Laplace transform of the adjusted forward process.

The next proposition characterizes the conditional Fourier Laplace transform of the log prices X_t and the adjusted forward process v_t through solutions of some Riccati equations.

Proposition 2.2.4. *Suppose that Assumption 2.1.2 holds, and let X be the process of log prices given by (2.1) and v be the adjusted forward process given by (2.23). Fix $T \geq 0$, $w \in \mathbb{C}$ with $\mathrm{Re}(w) \in [0,1]$ and $h \in \mathcal{G}_K^*$. Then the conditional Fourier-Laplace transform of (X, v)*

$$L_t(w, h; X_T, v_T) = \mathbb{E}\left[\exp\left(wX_T + \int_0^\infty h(\xi) v_T(\xi)\,\mathrm{d}\xi \right) \middle| \mathcal{F}_t \right], \quad t \leq T \tag{2.30}$$

can be computed thanks to the following formula

$$L_t(w, h; X_T, v_T) = \exp\left(w(X_t + r(T-t)) + \int_0^\infty \Psi(T-t, \xi; w, h) v_t(\xi)\,\mathrm{d}\xi \right), \tag{2.31}$$

where Ψ is the unique solution to the following Riccati equation

$$\begin{cases} \Psi(t, \xi; w, h) &= h(\xi - t)\mathbf{1}_{\{\xi \geq t\}} + \mathcal{R}\left(w, \int_0^\infty \Psi(t-\xi, z; w, h) K(z)\,\mathrm{d}z\right) \mathbf{1}_{\{\xi < t\}}, \\ \\ \Psi(0, \xi; w, h) &= h(\xi), \end{cases} \tag{2.32}$$

and the operator $\mathcal{R}$ is defined by

$$\mathcal{R}(w, \varphi) = \frac{1}{2}(w^2 - w) + \left(\rho\eta w - \lambda + \frac{\eta^2}{2}\varphi \right) \varphi. \tag{2.33}$$

Moreover,

$$\xi \to \Psi(t,\xi;w,h) \in \mathcal{G}_K^*, \quad t \geq 0, \tag{2.34}$$

and if

$$\mathrm{Re}(w) = 0, \quad \int_0^\infty \mathrm{Re}(h(\xi))v_T(\xi)\,\mathrm{d}\xi \leq 0$$

then

$$\int_0^\infty \mathrm{Re}(\Psi(T-t,\xi;w,h))v_t(\xi)\,\mathrm{d}\xi \leq 0, \quad t \leq T.$$

Remark 2.2.5. Existence and uniqueness of solutions to equations (2.32), satisfying (2.34), is proved in Section 2.5.1 (see Proposition 2.5.1). Notice that if Ψ solves (2.32) and we set

$$\psi(t) = \int_0^\infty \Psi(t,\xi;w,h)K(\xi)\,\mathrm{d}\xi, \tag{2.35}$$

then ψ solves the following *Volterra Riccati* equation

$$\psi(t) = \int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi + (K * \mathcal{R}(w,\psi(\cdot)))(t). \tag{2.36}$$

Conversely, if ψ solves (2.36) and we define

$$\Psi(t,\xi;w,h) = h(\xi-t)\mathbf{1}_{\{\xi\geq t\}} + \mathcal{R}(w,\psi(t-\xi))\,\mathbf{1}_{\{\xi<t\}} \tag{2.37}$$

then Ψ solves (2.32).

Proof of Proposition 2.2.4. Let Ψ the unique solution to (2.32), satisfying (2.34) (see Proposition 2.5.1). To simplify notation, throughout the proof we will omit the parameters w and h. Let Z be the semimartingale $Z_t = -\lambda\int_0^t V_s\,\mathrm{d}s + \eta\int_0^t \sqrt{V_s}\,\mathrm{d}W_s$, ψ be as in (2.35), and set

$$\theta = T-t, \quad \widetilde{Y}_t = \int_0^\infty \Psi(\theta,\xi)(v_t(\xi) - v_0(\xi+t))\,\mathrm{d}\xi.$$

The identity (2.24), equation (2.32), the stochastic Fubini theorem (see [Protter, 2005, Theorem 65]), and a change of variables yield

$$\begin{aligned}
\widetilde{Y}_t &= \int_0^\infty \int_0^t \Psi(\theta,\xi)K(t+\xi-s)\,\mathrm{d}Z_s\,\mathrm{d}\xi \\
&= \int_0^t \int_\theta^\infty h(\xi-\theta)K(t+\xi-s)\,\mathrm{d}\xi\,\mathrm{d}Z_s + \int_0^t \int_0^\theta \mathcal{R}(\psi(\theta-\xi))K(t+\xi-s)\,\mathrm{d}\xi\,\mathrm{d}Z_s \\
&= \int_0^t \int_{T-s}^\infty h(\xi-T+s)K(\xi)\,\mathrm{d}\xi\,\mathrm{d}Z_s + \int_0^t \int_{t-s}^{T-s} \mathcal{R}(\psi(T-s-\xi))K(\xi)\,\mathrm{d}\xi\,\mathrm{d}Z_s.
\end{aligned} \tag{2.38}$$

Equation (2.36) implies that

$$\psi(T-s) = \int_{T-s}^{\infty} h(\xi - T + s) K(\xi)\, \mathrm{d}\xi + \int_0^{T-s} \mathcal{R}(\psi(T-s-\xi)) K(\xi) d\xi. \tag{2.39}$$

We then plug (2.39) into (2.38) and obtain

$$\widetilde{Y}_t = \int_0^t \psi(T-s)\, \mathrm{d}Z_s - \int_0^t \int_0^{t-s} \mathcal{R}(\psi(T-s-\xi)) K(\xi)\, \mathrm{d}\xi\, \mathrm{d}Z_s. \tag{2.40}$$

We deduce, thanks to (2.40) and the stochastic Volterra equation for the variance process, the following semimartingale dynamics for the process $\widetilde{Y}$

$$\mathrm{d}\widetilde{Y}_t = \psi(\theta)\, \mathrm{d}Z_t - \mathcal{R}(\psi(\theta)) \int_0^t K(t-s)\, \mathrm{d}Z_s\, \mathrm{d}t = \psi(\theta)\, \mathrm{d}Z_t - \mathcal{R}(\psi(\theta))(V_t - v_0(t))\, \mathrm{d}t. \tag{2.41}$$

On the other hand, similar calculations show that

$$\int_0^{\infty} \Psi(\theta, \xi) v_0(\xi + t)\, \mathrm{d}\xi = \int_0^{\infty} h(\xi) v_0(\xi + T)\, \mathrm{d}\xi + \int_0^{\theta} \mathcal{R}(\psi(\xi)) v_0(T - \xi)\, \mathrm{d}\xi. \tag{2.42}$$

Define the process Y as

$$Y_t = \widetilde{Y}_t + \int_0^{\infty} \Psi(\theta, \xi) v_0(\xi + t)\, \mathrm{d}\xi = \int_0^{\infty} \Psi(\theta, \xi) v_t(\xi)\, \mathrm{d}\xi.$$

From equation (2.41) and (2.42) we obtain the following semimartingale dynamics for Y

$$dY_t = \psi(\theta)\, \mathrm{d}Z_t - \mathcal{R}(\psi(\theta)) V_t\, \mathrm{d}t. \tag{2.43}$$

Consider now the semimartingale

$$M_t = \exp(w(X_t - rt) + Y_t).$$

From equation (2.43) and Itô's formula, we obtain

$$\begin{aligned} \frac{\mathrm{d}M_t}{M_t} &= w\, \mathrm{d}X_t - wr\, \mathrm{d}t + \mathrm{d}Y_t + \frac{1}{2} w^2\, \mathrm{d}\langle X \rangle_t + \frac{1}{2}\, \mathrm{d}\langle Y \rangle_t + w\, \mathrm{d}\langle X, Y \rangle_t \\ &= -\frac{w}{2} V_t\, \mathrm{d}t + w\sqrt{V_t}\, \mathrm{d}B_t + \psi(\theta)\, \mathrm{d}Z_t - \mathcal{R}(\psi(\theta)) V_t\, \mathrm{d}t \\ &\quad + \frac{1}{2} w^2 V_t\, \mathrm{d}t + \frac{1}{2} \psi^2(\theta) \eta^2 V_t\, \mathrm{d}t + \rho \eta w \psi(\theta) V_t\, \mathrm{d}t \end{aligned} \tag{2.44}$$

where $B = \rho W + \sqrt{1-\rho^2} W^{\perp}$. From the definition of $\mathcal{R}$ in (2.33), we finally get

$$\frac{dM_t}{M_t} = w\sqrt{V_t}\, \mathrm{d}B_t + \psi(T-t)\eta\sqrt{V_t}\, \mathrm{d}W_t. \tag{2.45}$$

M is then a local martingale and

$$M_T = \exp\left(w(X_T - rT) + \int_0^{\infty} h(x) v_T(x) dx \right)$$

since $\Psi(0,\xi) = h(\xi)$. It is possible to show that ψ is a bounded function on $[0,T]$ thanks to the continuity of $\int_0^\infty h(\xi)K(\cdot+\xi)\,\mathrm{d}\xi$. Using a similar argument to the one used in Abi Jaber et al. [2019, Lemma 7.3], we can show that M is a true martingale. This implies the formula for the Fourier-Laplace transform (2.31). The last implication in the statement of the proposition is a direct consequence of (2.31).

□

To establish the convergence of approximated Bermudan option prices, we will use convergence results of the conditional Fourier-Laplace transform, which we present in the following section.

2.2.3 Convergence of the Fourier-Laplace transform

Suppose that the kernels $(K^n)_{n\geq 1}$ and the functions $(v_0^n)_{n\geq 1}$ satisfy Assumption 2.1.5. Let $(X^n, V^n)_{n\geq 1}$ be the solutions to (2.8) and let $(v^n)_{n\geq 1}$ be the corresponding adjusted forward processes as in (2.25). We define, analogously to (2.30), the associated conditional Fourier-Laplace transform

$$L^n(w, h^n; X_T^n, v_T^n) = \mathbb{E}^n\left[\exp\left(wX_T^n + \int_0^\infty h^n(\xi)v_T^n(\xi)\,\mathrm{d}\xi\right)\Big|\mathcal{F}_t^n\right], \tag{2.46}$$

where $h^n \in \mathcal{G}_{K^n}$ and $\mathrm{Re}(w) \in [0,1]$. Proposition (2.2.4) implies that

$$L_t^n(w, h^n; X_T^n, v_T^n) = \exp\left(wX_t^n + \int_0^\infty \Psi^n(T-t,\xi; w, h^n)v_t^n(\xi)\,\mathrm{d}\xi\right), \tag{2.47}$$

where Ψ^n solves (2.32) with h replaced by h^n and K replaced by K^n. We have the following convergence result for the conditional Fourier-Laplace transforms.

Proposition 2.2.6. *Suppose that Assumptions 2.1.2 and 2.1.5 hold. Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (2.23) (resp. (2.25)). Fix $T \geq 0$, $w \in \mathbb{C}$ with $\mathrm{Re}(w) \in [0,1]$, and $(h^n)_{n\geq 1}$ with $h^n \in \mathcal{G}^*_{K^n}$, $n \geq 1$. Assume that*

$$\mathrm{supp}(h^n) \subseteq [0,M],\ n \geq 1; \qquad \textit{and} \qquad h^n \to h \in \mathcal{G}^*_K \textit{ in } \mathcal{B}[0,M], \quad \textit{as } n \to \infty.$$

Then

$$L^n(w, h^n; X_T^n, v_T^n) \textit{ converges in law to } L(w, h; X_T, v_T) \textit{ in } \mathcal{C}[0,T], \quad \textit{as } n \to \infty,$$

where $L(w,h;X_T,v_T)$ and $L^n(w,h^n;X_T^n,v_T^n)$ are the conditional Fourier-Laplace transforms defined in (2.30) and (2.46), respectively.

The proof of Proposition 2.2.6 is based on Proposition 2.2.3 and the following lemma, whose proof is in Section 2.5.1.

Lemma 2.2.7. *Assume that the hypotheses of Proposition 2.2.6 hold. Let Ψ (resp. Ψ^n) be the solution to the Riccati equation (2.32) with kernel K (resp. K^n) and initial condition h (resp. h^n). Define*

$$\psi(t) = \int_0^\infty \Psi(t,\xi;w,h)K(\xi)\,\mathrm{d}\xi, \quad \psi^n(t) = \int_0^\infty \Psi^n(t,\xi;w,h^n)K^n(\xi)\,\mathrm{d}\xi.$$

Then, as n goes to infinity, ψ^n converges to ψ in $\mathcal{C}[0,T]$. Moreover, the support of $\Psi^n(t,\cdot;w,h^n)$ is contained in $[0,\max\{M,T\}]$ for all $n \geq 1$ and $t \geq 0$, and $\Psi^n(\cdot,\cdot;w,h^n)$ converges to $\Psi(\cdot,\cdot;w,h)$ in $\mathcal{C}([0,T]\times[0,M])$.

Proof of Proposition 2.2.6. By Proposition 2.2.3 and Skorohod's representation theorem we can construct (X^n,v^n) and (X,v) on the same probability space such that, as n goes to infinity, (X^n,v^n) converges almost surely to (X,v) in $\mathcal{C}(\mathbb{R}_+^2,\mathbb{R}^2)$. This observation and Lemma 2.2.7 imply that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\left|\omega X_t^n + \int_0^\infty \Psi^n(t,\xi;w,h^n)v_t^n(\xi)\,\mathrm{d}\xi - \omega X_t - \int_0^\infty \Psi(t,\xi;w,h)v_t(\xi)\,\mathrm{d}\xi\right| = 0, \quad \text{a.s.}$$

Hence, $X^n + \int_0^\infty \Psi^n(\cdot,\xi;w,h^n)v_t^n(\xi)\,\mathrm{d}\xi$ converges in law to $\omega X + \int_0^\infty \Psi(\cdot,\xi;w,h)v_t(\xi)\,\mathrm{d}\xi$ in $\mathcal{C}[0,T]$. An application of the continuous mapping theorem with the exponential function, together with Proposition 2.2.4, yields the conclusion. $\square$

We now possess all the elements necessary for the proof of Theorem 2.1.7.

2.3 Proof of the main convergence result

We break down the argument into different parts. We start by establishing, in the next section, the convergence of the Bermudan option prices as stated in (2.20). To this end, we will consider a more general payoff structure that is better suited for an inductive argument.

2.3.1 Convergence of Bermudan option prices

Throughout this section we will use the notation

$$\langle h,\hat{h}\rangle = \int_0^\infty h(\xi)\hat{h}(\xi)\,\mathrm{d}\xi,$$

for $h \in \mathcal{B}_c(\mathbb{R}_+, \mathbb{C})$ and $\hat{h} \in \mathcal{C}$. In addition, for a given finite set of indices J, we define

$$\mathcal{D}_J = \{(x, (\eta_j)_{j\in J}) \in (\mathbb{R}, \mathbb{C}^{\#J}) : \ \mathrm{Re}(\eta_j) \leq 0 \text{ for all } j \in J\}. \tag{2.48}$$

We will consider options with intrinsic payoff processes $(Z_t)_{0\leq t\leq T}$ defined as

$$Z_t = \begin{cases} f(X_t), & \text{for } 0 \leq t < T, \\ g(X_T, (\langle h_j, v_T\rangle)_{j\in J}), & \text{for } t = T, \end{cases} \tag{2.49}$$

where J is a finite set of indexes, $f \in \mathcal{C}_b(\mathbb{R})$, $g \in \mathcal{C}_b(\mathcal{D}_J)$, $h_j \in \mathcal{G}_K^*$ for all $j \in J$, and $(X_T, (\langle h_j, v_T\rangle)_{j\in J}) \in \mathcal{D}_J$. In this setting, the Bermudan option discrete value process over the grid $(t_i)_{i=0}^N$ takes the form

$$U_i^N = \operatorname{ess\,sup}_{\tau\in\mathcal{T}^N_{t_i,T}} \mathbb{E}\Big[\mathrm{e}^{-r(\tau-t_i)} Z_\tau | \mathcal{F}_{t_i}\Big], \quad 0 \leq i \leq N. \tag{2.50}$$

For the approximating models, and in an analogous manner, we will consider options with payoff processes $(Z_t^n)_{0\leq t\leq T}$ defined as

$$Z_t^n = \begin{cases} f(X_t^n), & \text{for } 0 \leq t < T, \\ g(X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J}), & \text{for } t = T, \end{cases} \tag{2.51}$$

where $h_j \in \mathcal{G}_{K^n}^*$, for all $j \in J$, and $(X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J}) \in \mathcal{D}_J$. The Bermudan option discrete value process, in the approximated model and over the grid $(t_i)_{i=0}^N$, takes the form

$$U_i^{N,n} = \operatorname{ess\,sup}_{\tau\in\mathcal{T}^N_{t_i,T}} \mathbb{E}^n\Big[\mathrm{e}^{-r(\tau-t_i)} Z_\tau^n | \mathcal{F}_{t_i}^n\Big], \quad 0 \leq i \leq N. \tag{2.52}$$

The following is the main result of this section.

Theorem 2.3.1. *Suppose that Assumptions 2.1.2 and 2.1.5 hold. Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (2.23) (resp. (2.25)). Fix $T \geq 0$, J a finite set of indexes, $f \in \mathcal{C}_b(\mathbb{R})$, $g \in \mathcal{C}_b(\mathcal{D}_J)$, and $(h^n)_{n\geq 1}$ with $h^n \in \mathcal{G}_{K^n}^*$, $n \geq 1$. Assume that*

$$\mathrm{supp}(h^n) \subseteq [0, M], \ n \geq 1; \quad \textit{and} \quad h^n \to h \in \mathcal{G}_K^* \textit{ in } \mathcal{B}[0, M] \quad \textit{as } n \to \infty.$$

Then

$$U_i^{N,n} \textit{ converges in law to } U_i^N, \quad i = 0, \ldots, N, \quad \textit{as } n \to \infty,$$

where U_i^N and $U_i^{N,n}$ are given by (2.50) and (2.52), respectively.

Proof. We prove the result by induction on the number of exercise dates $N+1$.

Initialization: Assume that $N=0$. We just have to prove that

$$\lim_{n\to+\infty} g(X_0, (\langle h_j^n, v_0^n\rangle)_{j\in J}) = g(X_0, (\langle h_j, v_0\rangle)_{j\in J}).$$

This follows from continuity of g on $\mathcal{D}_J$, because our hypotheses readily imply

$$\lim_{n\to+\infty} \langle h_j^n, v_0^n\rangle = \langle h_j, v_0\rangle.$$

Induction: Assume that the claim holds for Bermudan options with N exercise dates. We have to consider three different cases.

1) Suppose that g on $\mathcal{D}_J$ has the form

$$g(x, (\eta_j)_{j\in J}) = \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(i\left(\nu_k x + \sum_{j\in J}\beta_{j,k}\mathrm{Im}(\eta_j)\right) + \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(\eta_j)\right)\right), \quad (2.53)$$

with I a finite set of indices, $c_k\in\mathbb{C}$, $\nu_k\in\mathbb{R}$, $\alpha_{j,k}\geq 0$, $\beta_{j,k}\in\mathbb{R}$. In this case the value of the option at maturity (in the original Volterra model) is

$$Z_T = \mathrm{Re}\left(\sum_{k\in I} c_k \exp(i\nu_k X_T + \langle y_k(0), v_T\rangle)\right),$$

with

$$y_k(0) = \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(h_j) + i\sum_{j\in J}\beta_{j,k}\mathrm{Im}(h_j), \quad k\in I.$$

One can verify that for each $k\in I$, $y_k(0)\in\mathcal{G}_K^*$ thanks to the fact that $\alpha_{j,k}\geq 0$, $j\in J$, and the definition of $\mathcal{G}_K^*$ in (2.29) and $\mathcal{G}_K$ in (2.3). Applying the dynamic programming principle between t_{N-1} and T, we get

$$\begin{aligned} U_{N-1}^N &= \max\left(Z_{t_{N-1}},\ e^{-r(\Delta t_{N-1})}\mathbb{E}\left[U_T^N|\mathcal{F}_{t_{N-1}}\right]\right) \\ &= \max\left(f(X_{t_{N-1}}),\ e^{-r\Delta t_{N-1}}\mathbb{E}\left[g(X_T, (\langle h_j, v_T\rangle)_{j\in J}))|\mathcal{F}_{t_{N-1}}\right]\right), \end{aligned}$$

where $\Delta t_{N-1} = t_N - t_{N-1} = T - t_{N-1}$. According to the affine transform formula in Proposition 2.2.4, with w being purely imaginary, the value of the option at time $N-1$ is then

$$U_{N-1}^N = \max\left\{f(X_{t_{N-1}}), \mathrm{e}^{-r\Delta t_{N-1}}\mathrm{Re}\left(\sum_{k\in I} c_k \mathrm{e}^{i\nu_k(X_{t_{N-1}}+r\Delta t_{N-1})+\langle y_k(\Delta t_{N-1}), v_{t_{N-1}}\rangle}\right)\right\},$$

where $y_k(\Delta t_{N-1}) \in \mathcal{G}_K$ is the solution at time Δt_{N-1} of the associated Riccati equation (with initial condition $y_k(0)$), $k \in I$. Similarly, in the approximated model, we have

$$U_{N-1}^{N,n} = \max\left\{ f(X_{t_{N-1}}^n), \mathrm{e}^{-r\Delta t_{N-1}} \mathrm{Re}\left(\sum_{k\in I} c_k \mathrm{e}^{i\nu_k(X_{t_{N-1}}^n + r\Delta t_{N-1}) + \langle y_k^n(\Delta t_{N-1}), v_{t_{N-1}}^n \rangle} \right) \right\},$$

where $y_k^n(\Delta t_{N-1}) \in \mathcal{G}_{K^n}^*$ is the solution at time Δt_{N-1} of the associated Riccati equation with initial condition

$$y_k^n(0) = \sum_{j\in J} \alpha_{j,k} \mathrm{Re}(h_j^n) + i \sum_{j\in J} \beta_{j,k} \mathrm{Im}(h_j^n) \in \mathcal{G}_{K^n}^*.$$

Propositions 2.2.3 and 2.2.6 imply that $U_{N-1}^{N,n}$ converges in law to U_{N-1}^N. To prove that $U_i^{N,n}$ converges in law to U_i^N for $i = 0, \ldots, N-2$, we apply Lemma 2.2.7 together with the induction hypothesis in the case of a Bermudan option with maturity t_{N-1}, N exercise dates and final payoff $\hat{g}(X_{t_{N-1}}, (\langle \hat{h}_k, v_{t_{N-1}} \rangle)_{k\in I})$ where, for $k \in I$, $\hat{h}_k = y_k(\Delta t)$ and

$$\hat{g}(x, (\eta_k)_{k\in I}) = \max\left\{ f(x), , \mathrm{e}^{-r\Delta t_{N-1}} \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(i\nu_k(x + r\Delta t_{N-1}) + \eta_k \right) \right) \right\}.$$

Notice that $(X_{t_{N-1}}, (\langle \hat{h}_k, v_{t_{N-1}} \rangle)_{k\in I}) \in \mathcal{D}_I$ thanks to the last implication in Proposition 2.2.4.

2) Assume now that g vanishes outside a compact set $\Gamma \subset \mathcal{D}_J$.

Let $\varepsilon > 0$. By tightness of the sequence (X_T^n, v_T^n), its convergence to (X_T, v_T), and the convergence of h_j^n to h_j for all $j \in J$, there exists a compact set $\Gamma' \subset \mathcal{D}_J$ such that $\Gamma \subset \Gamma'$ and

$$\mathbb{P}\left(\left((X_T, (\langle h_j, v_T \rangle)_{j\in J}) \right) \notin \Gamma' \right) < \varepsilon, \quad \mathbb{P}^n\left(\left((X_T^n, (\langle h_j^n, v_T^n \rangle)_{j\in J}) \right) \notin \Gamma' \right) < \varepsilon, \quad n \geq 1. \tag{2.54}$$

Furthermore, we can assume that there exists a constant $A > 0$ such that

$$\Gamma' = \left\{ (x, (\eta_j)_{j\in J}) \in \mathcal{D}_J : \ |x| + \max_{j\in J}(|\eta_j|) \leq A \right\}.$$

Let $\mathcal{A}$ be an algebra of functions defined as follows. We say that a function $\hat{g}$ on $\mathcal{D}_J$ belongs to $\mathcal{A}$ if it is of the form

$$\hat{g}(x, (\eta_j)_{j\in J}) = \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(2\pi i \left(\frac{n_k}{2A} x + \sum_{j\in J} \frac{m_{k,j}}{2A} \mathrm{Im}(\eta_j) \right) + \sum_{j\in J} \alpha_{j,k} \mathrm{Re}(\eta_j) \right) \right),$$

with I a finite set of indices, $c_k \in \mathbb{C}$, $\alpha_{j,k} \geq 0$, and n_k and $m_{k,j}$ integers. We also define the following compact subset of $\mathcal{D}_J$

$$\widetilde{\Gamma} = \left\{(x, (\eta_j)_{j\in J}) \in \mathcal{D}_J : \ |x| + \max_{j\in J} |\mathrm{Im}(\eta_j)|) \leq A\right\}.$$

Notice that we have $\Gamma' \subset \widetilde{\Gamma}$ and, if we denote by $\mathcal{A}|_{\widetilde{\Gamma}}$ the restriction of all the functions in $\mathcal{A}$ to $\widetilde{\Gamma}$, $\mathcal{A}|_{\widetilde{\Gamma}}$ is a subset of $\mathcal{C}_0(\widetilde{\Gamma}, \mathbb{R})$ that satisfies the hypothesis of Stone-Weierstrass Theorem. Therefore, there exists $\hat{g} \in \mathcal{A}$ such that

$$\sup_{(x,(\eta_j)_{j\in J})\in\widetilde{\Gamma}} |g(x, (\eta_k)_{k\in I}) - \hat{g}(x, (\eta_k)_{k\in I})| \leq \varepsilon. \tag{2.55}$$

Now observe that for all $(x, (\eta_j)_{j\in J}) \in \mathcal{D}_J$, there exists $(x', (\eta'_j)_{j\in J}) \in \widetilde{\Gamma}$ such that $\hat{g}(x, (\eta_j)_{j\in J}) = \hat{g}(x', (\eta'_j)_{j\in J})$. Hence

$$\|\hat{g}\|_\infty \leq \varepsilon + \|g\|_\infty, \tag{2.56}$$

where $\|\cdot\|_\infty$ denotes the sup norm on $\mathcal{D}_J$.

Denote by $\widehat{U}^N$ (resp. $\widehat{U}^{N,n}$) the value processes for the Bermudan options corresponding to the payoff process $\widehat{Z}$ (resp. $\widehat{Z}^n$) obtained by replacing g by $\widehat{g}$ in (2.49) (resp. (2.51)). By the previous step we already know that

$$\widehat{U}_i^{N,n} \text{ converges in law to } \widehat{U}_i^N \text{ for } i = 1, \ldots, N-1. \tag{2.57}$$

Moreover, the dynamical programming principle yields

$$|U_i^N - \widehat{U}_i^N| \leq \mathbb{E}\left[|U_{i+1}^N - \widehat{U}_{i+1}^N|\Big|\mathcal{F}_{t_i}\right], \quad i = 0, \ldots, N-1.$$

By iterating this inequality, we deduce

$$|U_i^N - \widehat{U}_i^N| \leq \mathbb{E}\left[|g(X_T, \langle h_j, v_T\rangle)_{j\in J})) - \hat{g}(X_T, \langle h_j, v_T\rangle)_{j\in J}))||\mathcal{F}_{t_i}\right], \quad i = 0, \ldots, N.$$

Therefore, thanks to the inequalities (2.54), (2.55) and (2.56),

$$\mathbb{E}\Big[|U_i^N - \widehat{U}_i^N|\Big] \leq \varepsilon(1 + \|g\|_\infty + \|\hat{g}\|_\infty) \leq \varepsilon(1 + \varepsilon + \|g\|_\infty), \quad i = 0, \ldots, N. \tag{2.58}$$

Similarly we can prove that

$$\mathbb{E}^n\Big[|U_i^{N,n} - \widehat{U}_i^{N,n}|\Big] \leq \varepsilon(1 + \varepsilon + \|g\|_\infty), \quad i = 0, \ldots, N, \, n \geq 0. \tag{2.59}$$

Since ε is arbitrary we conclude, using (2.57), (2.58) and (2.59), that $U_i^{N,n}$ converges in law to U_i^N, for $i = 0, \ldots, N$.

3) Suppose now that g belongs to $\mathcal{C}_b(\mathcal{D}_J)$.

Let $\varepsilon > 0$ be arbitrary. As before, tightness of the sequence (X_T^n, v_T^n), its convergence to (X_T, v_T), and the convergence of h_j^n to h_j, $j \in J$, imply that there is a compact set $\Gamma \subset \mathcal{D}_J$ such that

$$\mathbb{P}\left(((X_T, (\langle h_j, v_T\rangle)_{j\in J})) \notin \Gamma)\right) < \varepsilon, \quad \mathbb{P}^n\left(((X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J})) \notin \Gamma)\right) < \varepsilon, \quad n \geq 1. \tag{2.60}$$

Let $\varphi : \mathcal{D}_J \to [0,1]$ be a function of compact support such that $\varphi \equiv 1$ on Γ.

Denote $\overline{U}^N$ (resp. $\overline{U}^{N,n}$) the value processes for the Bermudan options corresponding to the payoff process $\overline{Z}$ (resp. $\overline{Z}^n$) obtained by replacing g by $\overline{g} = \varphi g$ in (2.49) (resp. (2.51)). By the previous step we already know that

$$\overline{U}_i^{N,n} \text{ converges in law to } \overline{U}_i^N \text{ for } i = 1, \ldots, N-1. \tag{2.61}$$

Additionally, we have

$$\begin{aligned} \mathbb{E}\Big[|U_i^N - \overline{U}_i^N|\Big] &\leq \mathbb{E}\left[|g(X_T, \langle h_j, v_T\rangle)_{j\in J})) - \overline{g}(X_T, \langle h_j, v_T\rangle)_{j\in J}))|\right] \\ &\leq \varepsilon \|g\|_\infty, \end{aligned} \tag{2.62}$$

and

$$\mathbb{E}^n\Big[|U_i^{N,n} - \overline{U}_i^{N,n}|\Big] \leq \varepsilon \|g\|_\infty. \tag{2.63}$$

Since ε is arbitrary we conclude, from (2.61), (2.62) and (2.63), that $U_i^{N,n}$ converges in law to U_i^N, for $i = 0, \ldots, N$.

$\square$

2.3.2 Approximation of American options with Bermudan options

The following theorem establishes the convergence of Bermudan option prices towards American option prices and it is crucial in order to prove (2.21) in Theorem 2.1.7.

Theorem 2.3.2. *Suppose that Assumption 2.1.2 holds. Let (X, V) be the unique weak solution to (2.1). For a function $f \in \mathcal{C}_b^2(\mathbb{R})$ consider the American and Bermudan option prices given by (2.7) and (2.17), respectively. Then*

$$0 \leq P_0 - P_0^N \leq c\left(1 + \mathbb{E}\Big[\sup_{t\in[0,T]} V_t\Big]\right)\pi_N, \tag{2.64}$$

where π_N is the mesh of the partition $(t_i)_{i=0}^N$ and c is a constant that only depends on r, T and $\|f^{(m)}\|_{\mathcal{C}[0,T]}$, $m = 0, 1, 2$.

Proof. We obviously have $0 \leq P_0 - P_0^N$. Let $\varepsilon > 0$. There exists $\tau_\varepsilon^* \in \mathcal{T}_{0,T}$, ε-optimal in the sense that

$$P_0 \leq \mathbb{E}\Big[\mathrm{e}^{-r\tau_\varepsilon^*} f(X_{\tau_\varepsilon^*})\Big] + \varepsilon.$$

Now, we introduce the lowest stopping time taking values in $\{t_0, ..., t_N\}$, greater than τ_ε^*, this is

$$\tau_\varepsilon^{N,*} = \inf\{t_k : \; t_k \geq \tau_\varepsilon^*\}.$$

We have that $\tau_\varepsilon^{N,*}$ belongs to $\mathcal{T}_{0,T}^N$. Applying Itô's formula to the process $\big(\mathrm{e}^{-rt} f(X_t)\big)_{0\leq t\leq T}$ between τ_ε^* and $\tau_\varepsilon^{N,*}$ yields

$$\begin{aligned} P_0 - P_0^N \quad &\leq \quad c\mathbb{E}\Big[\int_{\tau_\varepsilon^*}^{\tau_\varepsilon^{N,*}} (1 + V_s)\, ds\Big] + \varepsilon &(2.65)\\ &\leq \quad c\mathbb{E}\Big[(\tau_\varepsilon^{N,*} - \tau_\varepsilon^*) \sup_{t\in[0,T]} (1 + V_t)\Big] + \varepsilon &(2.66)\\ &\leq \quad c\left(1 + \mathbb{E}\Big[\sup_{t\in[0,T]} V_t\Big]\right) \pi_N + \varepsilon, &(2.67)\end{aligned}$$

where c is a constant that only depends on r, T and $\|f^{(m)}\|_{\mathcal{C}[0,T]}$, $m = 0, 1, 2$. Since $\varepsilon > 0$ was arbitrary, we deduce (2.64). □

We are now ready to prove our main theorem.

Proof of Theorem 2.1.7. The convergence in (2.20) is a direct consequence of Theorem 2.3.1. On the other hand, (2.27) and Theorem 2.3.2 yield (2.21). The limit (2.22) follows from (2.20) and (2.21). □

2.4 Numerical illustrations

In this section we illustrate with numerical examples the convergence and behavior of Bermudan put option prices in the approximated sequence of models. To this end, we consider the framework of the rough Heston model in Example 2.1.4 and the approximation scheme of Example 2.1.6.

We choose the same model parameters as in Abi Jaber [2019a], namely

$$V_0 = 0.02, \quad \bar{\nu} = 0.02, \quad \lambda = 0.3, \quad \eta = 0.3, \quad \rho = -0.7. \tag{2.68}$$

We fix a maturity $T = 0.5$ and a spot interest rate $r = 0.06$.

In order to calculate Bermudan option prices in the approximated model (X^n, V^n) in (2.8), we apply the Longstaff Schwartz algorithm [Longstaff and Schwartz, 2001] using 10^5 path simulations. Following the suggestion in Abi Jaber [2019a], and based on the factor-representation (2.14), we simulate the trajectories of the variance process with a *truncated* explicit-implicit Euler-scheme and the trajectories of the log prices with an explicit Euler-scheme. More precisely, given a uniform partition $(s_k)_{k=0}^{N_{time}}$ of $[0,T]$ of norm Δt, and $\left(G_1^k\right)_{k\geq 1}$ and $\left(G_2^k\right)_{k\geq 1}$ independent sequences of independent centered and reduced gaussian variables, we simulate the log prices with the scheme

$$\widehat{X}^n_{s_{k+1}} = \widehat{X}^n_{s_k} + \left(r - \frac{\widehat{V}^n_{s_k}}{2}\right)\Delta t + \sqrt{\widehat{V}^{n\ +}_{s_k}}\sqrt{\Delta t}\left(\rho G_1^{k+1} + \sqrt{1-\rho^2}G_2^{k+1}\right), \quad \widehat{X}^n_{s_0} = X_0,$$

and the variance process with the scheme

$$\widehat{V}^n_{s_k} = v_0^n(s_k) + \sum_{i=1}^n c_i^n \widehat{U}^{n,i}_{s_k}, \quad \widehat{U}^{n,i}_0 = 0, \quad i = 1,\dots,n,$$

$$\widehat{U}^{n,i}_{s_{k+1}} = \frac{1}{1+x_i^n\Delta t}\left(\widehat{U}^{n,i}_{s_k} - \lambda \widehat{V}^n_{s_k}\Delta t + \eta\sqrt{\widehat{V}^{n\ +}_{s_k}}\sqrt{\Delta t}G_1^{k+1}\right), \quad i = 1,\dots,n.$$

In this framework the initial curve v_0 in (2.6) takes the form

$$v_0^n(s_k) = V_0 + \lambda\bar{\nu}\sum_{i=1}^n c_i^n\left(\frac{1-e^{-x_i^n s_k}}{x_i^n}\right).$$

We take $N_{time} = 500$ and select equidistant exercise times $(t_k)_{i=0}^N$, with $N = 50$, within the partition $(s_k)_{k=0}^{N_{time}}$. Given a strike price K, for the regressions of the Longstaff Schwartz algorithm we use the linear space of functions generated by functions with argument S, corresponding to the log price, and V corresponding to the volatility, of the form

$$f_1\left(\frac{S}{K}\right)f_2\left(\frac{V}{\bar{\nu}}\right), \quad f_1, f_2 \in \mathcal{A}$$

where $\mathcal{A}$ is given by

$$\mathcal{A} = \{1\} \cup \{\mathrm{e}^{-z}L_i(z) : i = 0,1,2\},$$

and L_i denotes the Laguerre polynomial of order i.[5]

[5] In the framework of our factor-approximation scheme, the prices of the Bermudan options at intermediate times are functions of the log price S and the factors $(Y^{n,i})_{i=1}^n$ defined in (2.14). This functions could be approximated using neural network-based techniques similar to those in Lapeyre and Lelong [2019]. Our initial experiments, however, indicate that there is no significant gain in using this more complex approach. This is consistent with similar findings in Bayer et al. [2020b] and Goudenège et al. [2020] for American options prices in the rough Bergomi model.

To illustrate the convergence of options prices, we fix the parameter $\alpha = 0.6$ and choose parameters $r_n > 1$ in the kernel approximation such that

$$\begin{aligned} r_n &= \arg\min_r \|K - K^r\|^2_{\mathcal{L}^2(0,T)} \\ &= \arg\min_r \left(\sum_{i,j\le n} c_i^r c_j^r \frac{1 - \mathrm{e}^{-(x_i^r + x_j^r)T}}{x_i^r + x_j^r} - 2\sum_{i\le n} c_i^r (x_i^r)^{-\alpha} \gamma(\alpha, T x_i^r) \right), \end{aligned} \tag{2.69}$$

where c_i^r, x_i^r, $i = 1, \ldots, n$, are as in (2.16) with r_n replaced by r, K^r is the corresponding kernel obtained as a sum of exponentials, and $\gamma(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} \mathrm{e}^{-t}\, \mathrm{d}t$ is the lower incomplete gamma function. Table 2.1 contains the values of the parameter r_n along with the corresponding values of $\|K - K^n\|^2_{\mathcal{L}^2(0,T)}$ for $n = 4, 10, 20, 40, 200$.

n	r_n	norm_n^2
4	50.5458	0.3699
10	18.0548	0.1125
20	8.8750	0.0325
40	4.4737	0.0076
200	1.6946	1.1166e-04

Table 2.1: Values of r_n and $\text{norm}_n^2 = \|K - K^n\|^2_{\mathcal{L}^2(0,T)}$ obtained using (2.69) with $\alpha = 0.6$ and $T = 0.5$.

Figure 2.1 shows Bermudan put option prices for a strike $K = 100$, initial prices $S_0 = \exp(X_0)$ in $[93, 96]$, and $n = 4, 10, 20, 40$ number of factors. We also plot the prices obtained for the classical Heston model. For each set of prices we indicate the corresponding so-called critical price, this is the greatest value of the initial price for which the Bermudan option price is equal to the payoff. We observe that as n increases the option prices on this interval decrease and as a result the critical price increases. In Figure 2.2, we plot the critical-price as a function of the norm $\|K - K^n\|_{\mathcal{L}^2(0,T)}$ for $n = 1, 4, 10, 20, 40$, where $n = 1$ corresponds to the classical Heston model. Computing prices with $n = 200$ factors we observe the same critical price as with $n = 40$ which illustrates the convergence of the approximated models.

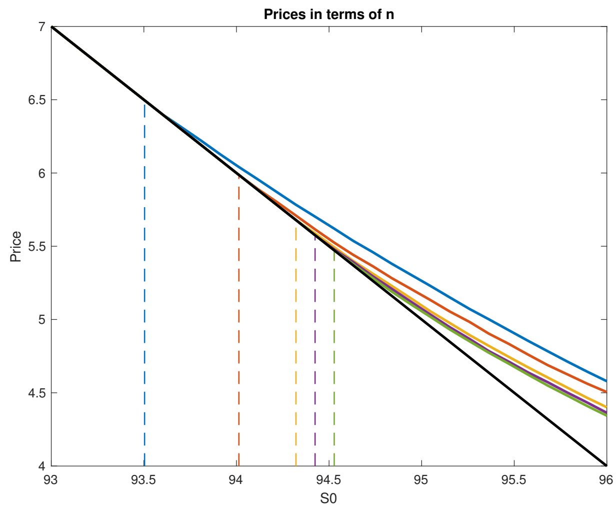

Figure 2.1: Bermudan put option prices in terms of n. Payoff (solid black), Heston model (blue), $n = 4$ (red), $n = 10$ (yellow), $n = 20$ (purple), $n = 40$ (green).

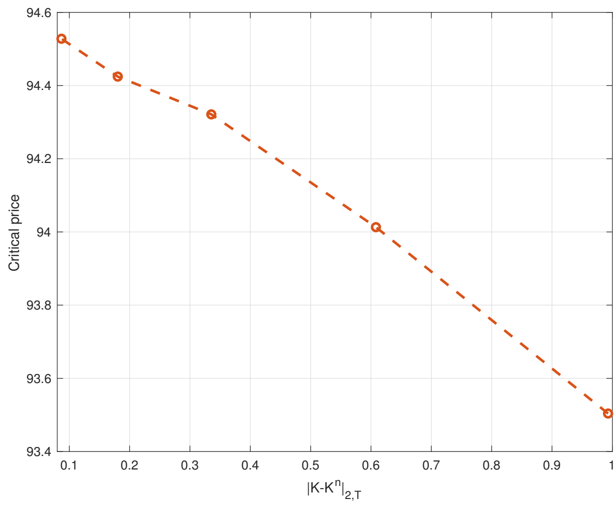

Figure 2.2: Critical prices as a function of $\|K - K^n\|_{\mathcal{L}^2(0,T)}$.

To study the behavior of Bermudan put option prices with respect to the parameter α, and taking into account our previous findings, we proxy the prices in the rough Heston model using the approximated model with $n = 40$ factors. We consider the same parameters as in the previous example with the exception of α. The parameter r_{40} is chosen as in (2.69) depending on the parameter α of the fractional kernel K. We compute prices and critical prices for $\alpha = 0.6, 0.7, 0.8, 0.9, 1$. Figure 2.3 shows the Bermudan option prices obtained for these values of α and Figure 2.4 displays the critical price as a function of α. As α increases, we observe a similar behavior as the one obtained by increasing $\|K - K^n\|_{\mathcal{L}^2(0,T)}$ in our previous example. More precisely, as the regularity of the paths in the model increases, i.e. α increases, the prices of the option increase and the critical price decreases. This is consistent with similar findings reported in Horvath et al. [2017] within the context of the rough Bergomi model and it could be a consequence of the fact that for smaller values of α the variance process has rougher paths and spends more time in a small neighborhood of zero.

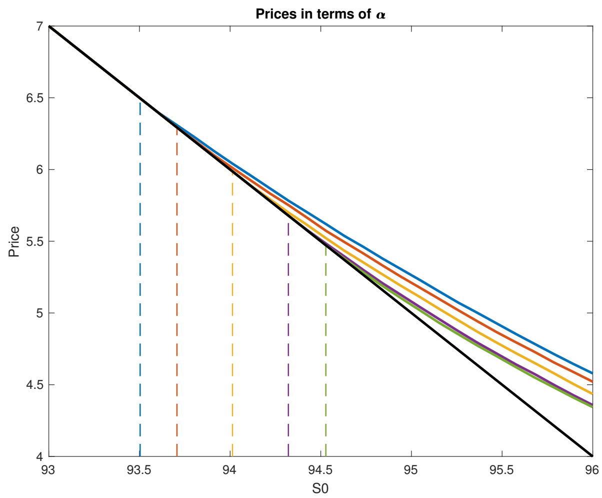

Figure 2.3: Bermudan put option prices and critical prices in terms of α. Payoff (black), $\alpha = 1$ (blue), $\alpha = 0.9$ (red), $\alpha = 0.8$ (yellow), $\alpha = 0.7$ (purple), $\alpha = 0.6$ (green).

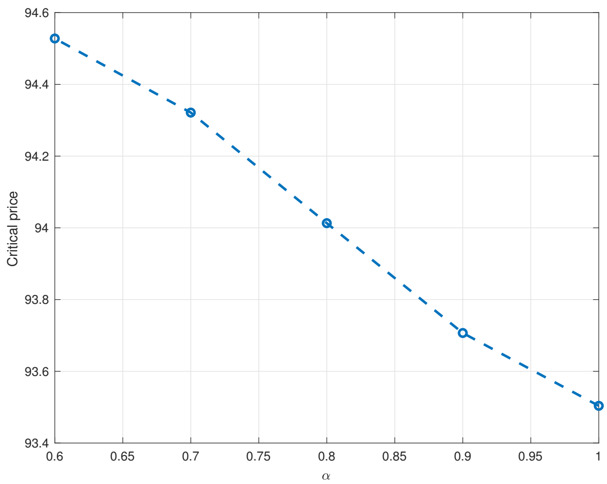

Figure 2.4: Critical prices as a function of α.

To illustrate the impact of the initial spot variance, we compare in Figure 2.5 the levels of the critical price for different values of V_0 in the rough Heston model with $\alpha = 0.6$ and the classical Heston model. As already pointed out in , the critical price seems to depend almost linearly on the initial spot variance V_0 in both the classical and the rough Heston model. In the rough Heston model the critical price, and hence the Bermudan option prices, seem to be slightly less sensitive to the initial level of the variance. This could be a result of the difference in sensitivity, with respect to V_0, of the time spent around zero by the trajectories in the classical and rough Heston models.

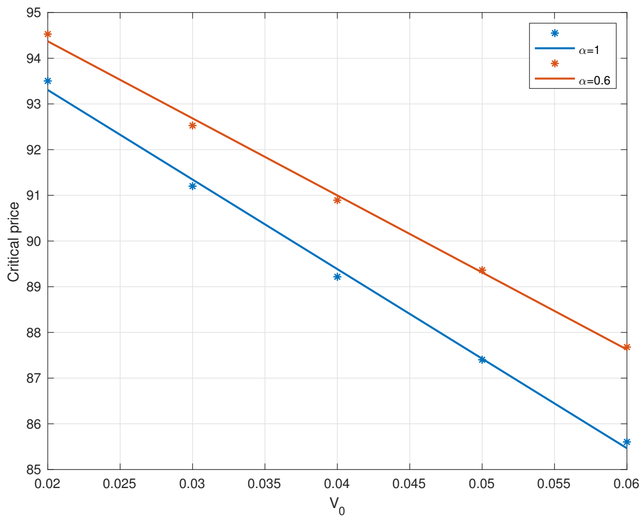

Figure 2.5: Critical prices for $\alpha = 0.6, 1$ and $V_0 = 0.02 + k * 0.01$, $k = 0, 1, 2, 3, 4$. The solid lines represent the linear regressions.

A theoretical explanation of our numerical findings would require more detailed results about the path-behavior of the rough Heston model, and their impact on American and Bermudan option prices. Such study falls outside of the scope of this manuscript and it could be an interesting topic of future research, along with a deeper numerical analysis of the behavior of American and Bermudan option prices in terms of the parameters of rough volatility models.

2.5 Support Results

2.5.1 Volterra Riccati equations

Proposition 2.5.1. *Suppose that $K \in \mathcal{L}^2_{loc}$ satisfies condition 1 in Assumption 2.1.2. Then, given $w \in \mathbb{C}$ with $\mathrm{Re(w)} \in [0,1]$, and $h \in \mathcal{G}^*_k$, the Riccati Volterra equation (2.32) admits a unique solution Ψ on $\mathbb{R}^2_+$ such that $\Psi(t,\cdot;w,h) \in \mathcal{G}^*_K$ for all $t \geq 0$.*

Proof. As pointed out in Remark 2.2.5 the Riccati equation (2.32) for Ψ can be recast as the stochastic Volterra equation (2.36) for the function ψ defined by (2.35). The existence and uniqueness of an $\mathcal{L}^2$-solution ψ on a maximal interval $[0, T_{max})$ is guaranteed by Abi Jaber et al. [2019, Theorem B.1]. In our setting, the function ψ can be shown to be bounded on $[0, T_{max})$ thanks to the continuity of $\int_0^\infty h(\xi)K(\cdot+\xi)\,\mathrm{d}\xi$. This in turn implies the continuity of ψ on $[0, T_{max})$. In Abi Jaber et al. [2019] the authors consider constant initial conditions, in our case we do not have a constant initial condition but instead an initial curve $\int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi$ such that $-\int_0^\infty \mathrm{Re}(h(\xi))K(t+\xi)\,\mathrm{d}\xi \in \mathcal{G}_K$. In order to prove that $T_{max} = \infty$ we can, however, follow the proof of Abi Jaber et al. [2019, Lemma 7.4], using the invariance result in Abi Jaber and El Euch [2019a, Theorem C.1] together with the fact that $\int_0^\infty f(\xi)K(t+\xi)\,\mathrm{d}\xi \in \mathcal{G}_K$ for all real-valued non-negative functions $f \in \mathcal{B}$. Moreover, by taking minus the real part in (2.36), Abi Jaber and El Euch [2019a, Theorem C.1] guarantees that

$$g_t(s) = \Delta_t g(s) - (\Delta_s K * \mathrm{Re}(\mathcal{R}(\omega,\psi)))(t) \in \mathcal{G}_K, \quad t,s \leq T.$$

with $g(s) = -\int_0^\infty \mathrm{Re}(h(\xi))K(s+\xi)\,\mathrm{d}\xi$. The last part of the statement is a consequence of the identity

$$\Delta_t g(s) - (\Delta_s K * \mathrm{Re}(\mathcal{R}(\omega,\psi)))(t) = -\int_0^\infty \mathrm{Re}(\Psi(t,\xi;w,h))K(s+\xi)\,\mathrm{d}\xi.$$

□

We finish this section with a sketch of the proof of Lemma 2.2.7.

Proof of Lemma 2.2.7. To prove the convergence of ψ^n towards ψ in $\mathcal{C}[0,T]$, one can use similar arguments as in the proof of Abi Jaber and El Euch [2019a, Theorem 4.1], replacing the zero initial condition by the initial curves $\int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi$ and $\int_0^\infty h^n(\xi)K^n(t+\xi)\,\mathrm{d}\xi$, $n \geq 1$. The convergence of Ψ^n towards Ψ is a consequence of the identity (2.37), the convergence of (h^n, ψ^n) to (h,ψ), and the quadratic structure of $\mathcal{R}(w,\cdot)$. Since $\mathrm{supp}(h^n) \subseteq [0,M]$ for all $n \geq 1$, thanks to the form of the Riccati equations satisfied by Ψ^n, we conclude that $\operatorname{supp}(\Psi^n(t,\cdot;w,h^n)) \subseteq [0,\max\{T,M\}]$ for all $n \geq 1$ and $t \geq 0$. $\square$

2.5.2 Some results on the kernel approximation

In this section we provide sufficient conditions on the kernel approximation which ensure condition 1 in Assumption 2.1.5.

Theorem 2.5.2. *Suppose that μ is a non-negative Borel measure on $\mathbb{R}_+$ such that*

$$\int_{\mathbb{R}_+} (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu(\mathrm{d}x) \leq c(T)\varepsilon^{\frac{\gamma-1}{2}}, \quad T > 0,\, \varepsilon \leq T^{6} \tag{2.70}$$

with $\gamma \in (0,2]$ and $c : \mathbb{R}_+ \to \mathbb{R}_+$ a locally bounded function. If in addition

$$\sup_{n\geq 1} \sup_{i\in\{1,\dots n\}} \frac{\eta^n_{i+1}}{\eta^n_i} < \infty$$

then the kernels $(K_n)_{n\geq 1}$ defined in (2.13), with $(c^n_i)^n_{i=1}$, $(x^n_i)^n_{i=1}$ given by (2.15), satisfy condition 1 in Assumption 2.1.5 with γ as in (2.70).

To prove Theorem 2.5.2 we use the following lemma.

Lemma 2.5.3. *Suppose that μ is a non-negative Borel measure μ on $\mathbb{R}_+$ such that (2.70) holds. Let K be the corresponding completely monotone kernel as in (2.11). Then K satisfies condition 1 in Definition 2.1.1, with the locally bounded function $2c^2$ and the same constant γ as in (2.70).*

Proof. Note that

$$\|K\|_{\mathcal{L}^2(0,\varepsilon)} \leq \int_0^\infty \|\mathrm{e}^{-\cdot x}\|_{\mathcal{L}^2(0,\varepsilon)}\mu(\mathrm{d}x) = \int_0^\infty \sqrt{\frac{1-\mathrm{e}^{-2x\varepsilon}}{2x}}\mu(\mathrm{d}x) \leq \varepsilon^{\frac{1}{2}} \int_0^\infty (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu(\mathrm{d}x),$$

This implies, by (2.70), that $\|K\|_{\mathcal{L}^2(0,\varepsilon)} \leq c(T)\epsilon^{\frac{\gamma}{2}}$, $\varepsilon \leq T$. A similar argument shows that $\|\Delta_\varepsilon K - K\|_{\mathcal{L}^2(0,T)} \leq c(T)\epsilon^{\frac{\gamma}{2}}$. As a result, condition 1 in Definition 2.1.1 holds with the locally bounded function $2c^2$ and the same constant γ as in (2.70). $\square$

Proof of 2.5.2. According to Lemma 2.5.3 it is enough to show that there is a locally bounded function $\tilde{c} : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\int_{\mathbb{R}_+} (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu^n(\mathrm{d}x) \leq \tilde{c}(T)\varepsilon^{\frac{\gamma-1}{2}}, \quad T > 0,\, \varepsilon \leq T, n \geq 1,$$

where μ^n is a sum of Dirac measures as in (2.12). This is a routine verification, using (2.70) and the definition of c_i, x_i in (2.15), and we omit the details. $\square$

[6] This condition was considered also in Abi Jaber and El Euch [2019a, Section 4].

Remark 2.5.4. Let K be the fractional kernel (2.5) and consider the geometric partition $\eta_i^n = r_n^{i-\frac{n}{2}}$, $i = 0, \ldots, n$. It is easy to check that the hypotheses of Theorem 2.5.2 hold with $\gamma = 2\alpha - 1$ as long as $\sup_{n \geq 1} r_n < \infty$.

Chapter 3

Moments of Polynomial Volterra processes

This chapter is organized as follows. The first section 3.1 is devoted to a brief introduction of polynomial diffusions. These processes have a special property, their moments are polynomial functions of their initial state and can be computed through the evaluation of an exponential matrix. Affine diffusions are a particular case of polynomial diffusions. In Section 3.2 we give a formula to compute moments in a polynomial Volterra setting. In Section 3.3 we show that the moments of affine Volterra processes are polynomial functions of their initial state, as in classical affine diffusions. Also, we present some general formulas to calculate the conditional moments of affine Volterra processes. At the end of the chapter, in Section 3.4, we present the numerical implementation of the moments. We use a multi-factor approximation of affine Volterra processes. This approximation enjoys of the properties of affine diffusions, so that its moments can be computed as in the classical case, using an exponential matrix.

Notation

A polynomial p on $\mathbb{R}^d$ is a map $\mathbb{R}^d \to \mathbb{R}$ of the form $\sum_\alpha \tilde{c}_\alpha x_1^{\alpha_1} \dots x_d^{\alpha_d}$, where the sum runs over all multi-indices $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$. Such a representation is unique. The degree of p is the number $deg \quad p = \max\{\alpha_1 + \dots + \alpha_d : c_\alpha \neq 0\}$. Let $\mathrm{Pol}(\mathbb{R}^d)$ denote the ring of all polynomials on $\mathbb{R}^d$ and $\mathrm{Pol}_n(\mathbb{R}^d)$ the subspace consisting of polynomials of degree at most n. Let E be a subset of $\mathbb{R}^d$, a polynomial on E is the restriction $p = q|_E$ to E of a polynomial $q \in \mathrm{Pol}(\mathbb{R}^d)$ with degree $deg \quad p = \min\{deg \quad q : p = q|_E, q \in \mathrm{Pol}(\mathbb{R}^d)\}$. We let $\mathrm{Pol}(E)$ be the ring of polynomials

on E and $\mathrm{Pol}_n(E)$ the subspace of polynomials on E of degree at most n.
Given a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ we denote its norm by $|\alpha| = \sum_{i=1}^d \alpha_i$. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and a vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ we write $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$. Finally, we let $\mathbb{I}_d$ be the identity matrix of size $d \times d$.

3.1 Polynomial Diffusions

We are going to study polynomial diffusions as introduced in Filipović and Larsson [2016]. Let X satisfy the following stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad t \geq 0, \tag{3.1}$$

where W is a d-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R} \to \mathbb{R}^{d\times d}$. Here, we assume X_0 deterministic. The functions b and $a = \sigma\sigma^\top : \mathbb{R}^d \to \mathbb{S}^d$ have the following form,

$$a_{ij} \in \mathrm{Pol}_2(\mathbb{R}^d) \text{ and } b_i \in \mathrm{Pol}_1(\mathbb{R}^d) \text{ for all } i, j. \tag{3.2}$$

where $\mathbb{S}^d$ is the set of real symmetric $d \times d$ matrices. Let $\mathcal{G}$ be the partial differential operator given by:

$$\mathcal{G}f = \frac{1}{2}\mathrm{Tr}(a\nabla^2 f) + b^T \nabla f, \tag{3.3}$$

for $f \in \mathrm{Pol}(\mathbb{R}^d)$. Assume in the following that the state space of X is $E \subseteq \mathbb{R}^d$.

Definition 3.1.1. The operator $\mathcal{G}$ in (3.3) is called polynomial on E if $\mathcal{G}\mathrm{Pol}_n(E) \subseteq \mathrm{Pol}_n(E)$. In this case, we call any E-valued solution X to (3.1) a *polynomial diffusion* on E.

For any $n \in \mathbb{N}$, we denote by $N = N(n, E)$ the dimension of $\mathrm{Pol}_n(E)$. We fix a basis of polynomials $h_1, \dots, h_N$ for $\mathrm{Pol}_n(E)$ and define,

$$H(x) = (h_1(x), \dots, h_N(x))^\top.$$

This means that for each $p \in \mathrm{Pol}_n(E)$ there exists a unique vector $\vec{p} \in \mathbb{R}^N$ such that, $p(x) = H(x)^\top \vec{p}$.

If $\mathcal{G}$ is polynomial, then for $p \in \mathrm{Pol}_n(E)$, we have:

$$\mathcal{G}p(x) = H(x)^\top G\vec{p},$$

where the matrix $G \in \mathbb{R}^{N\times N}$ is the matrix of the operator $\mathcal{G}$ with respect to H.

The following theorem shows that $\mathbb{E}[p(X_T)|\mathcal{F}_t]$ is a polynomial function of X_t. In particular, if X is a polynomial diffusion, then the moments of X_t are polynomial functions of the initial state X_0.

Theorem 3.1.2. *Let X be a polynomial diffusion on E. Then, for any $p \in Pol_n(E)$ with coordinate representation $\vec{p} \in \mathbb{R}^N$, we have*

$$\mathbb{E}[p(X_T)|\mathcal{F}_t] = H(X_t)^\top e^{(T-t)G}\vec{p}, \qquad t \leq T.$$

Example 3.1.3. Consider the following process in $\mathbb{R}$:

$$dX_t = (b_0 + b_1 X_t)dt + \sqrt{a_0 + a_1 X_t + a_2 X_t^2} dW_t \quad t \geq 0,$$

with b_0, b_1, a_0, a_1 and a_2 in $\mathbb{R}$.

The partial differential operator is given by:

$$\mathcal{G}f(x) = \frac{1}{2}\left(a_0 + a_1 x + a_2 x^2\right)\frac{\partial^2 f(x)}{\partial x^2} + (b_0 + b_1 x)\frac{\partial f(x)}{\partial x}.$$

Let $H(x) = (x^0, x^1, \ldots, x^m)$. Applying $\mathcal{G}$ to H yields the following $(m+1)\times(m+1)$ matrix:

$$G = \begin{pmatrix} 0 & b_0 & a_0 & 0 & \ldots & 0 \\ \vdots & b_1 & 2b_0 + a_1 & 3a_0 & 0 & \vdots \\ & 0 & 2b_1 + a_2 & 3b_0 + 3a_1 & & 0 \\ & \vdots & 0 & 3b_1 + 3a_2 & \ddots & m\frac{a_0}{2}(m-1) \\ & & \vdots & 0 & \ddots & m\left(b_0 + \frac{a_1}{2}(m-1)\right) \\ 0 & & \ldots & & 0 & m\left(b_1 + \frac{a_2}{2}(m-1)\right) \end{pmatrix}$$

Hence,

$$\mathbb{E}\left[(X_t)^k | X_0 = x\right] = H(x)^\top e^{tG}\vec{p},$$

where $\vec{p} = (0, \ldots, 1, \ldots, 0)^\top$ is the column vector with the value 1 in the position k.

3.2 Computation of moments of Polynomial Volterra processes

We consider the stochastic Volterra equation (1.25) in $\mathbb{R}$, that is:

$$X_t = u_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \geq 0, \tag{3.4}$$

where W is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $u_0 : \mathbb{R}_+ \to \mathbb{R}$ is a real-valued continuous function, $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$, and coefficients $b, \sigma : \mathbb{R} \to \mathbb{R}$ continuous. Also, we consider an affine drift and a quadratic diffusion:

$$b(x) = b_0 + b_1 x \quad \text{and} \quad a(x) = \sigma(x)^2 = a_0 + a_1 x + a_2 x^2, \tag{3.5}$$

with b_0, b_1, a_0, a_1, $a_2 \in \mathbb{R}$. Let $E \subseteq \mathbb{R}$ be the state space. We refer to the continuous E-valued solutions of (3.4) with coefficients (3.5) as polynomial Volterra processes. Theorem 1.4.1 gives the conditions for the existence of weak solutions for these processes.

The following theorem studied in Abi Jaber et al. [2021b] gives a way to compute the moments of X in (3.4) and its proof makes use of the adjusted forward process studied in Section 1.4.3. We recall the definition of the adjusted forward process, which is given by:

$$u_t(x) \quad = \quad \mathbb{E}\left[X_{t+x} - \int_0^x K(x-s) b(X_{t+s}) ds \middle| \mathcal{F}_t \right], \qquad t, x \geq 0. \tag{3.6}$$

Notice that $u_t(0) = X_t$. Based on the adjusted forward, we define the process m_t^p as follows:

$$m_t^p(x) = \mathbb{E}\left[\prod_{i=1}^p u_t(x_i) \right], \qquad t, x \geq 0. \tag{3.7}$$

The moments of order p of the process X satisfying (3.4) are given by evaluating m_t^p at $x = 0$, because for $t \geq 0$:

$$m_t^p(0) = \mathbb{E}\left[\prod_{i=1}^p u_t(0) \right] = \mathbb{E}[X_t^p].$$

Theorem 3.2.1. *Let X satisfy the stochastic Volterra equation (3.4). Then, m_t^p in (3.7) satisfies the following equation:*

$$\begin{aligned} m_t^p(x) = \prod_{i=1}^p u_0(x_i+t) + \sum_{i=1}^p \int_0^t K(t-s+x_i) \big(b_0 m_s^{p-1}((t-s+x_j)_{j\neq i}) + b_1 m_s^p\big((t-s+x_j)_{j\neq i}, 0\big)\big) ds \\ + \sum_{1\leq i<j\leq p} \int_0^t K(t-s+x_i) K(t-s+x_j) \big(a_0 m_s^{p-2}((t-s+x_k)_{k\neq i,j}) + a_1 m_s^{p-1}((t-s+x_k)_{k\neq i,j}, 0) + \\ a_2 m_s^p((t-s+x_k)_{k\neq i,j}, 0, 0)\big) ds, \end{aligned} \tag{3.8}$$

which is the mild formulation of the following PDE:

$$\begin{aligned} \partial_t m_t^p(x) = (\partial_1 + \ldots + \partial_p) m_t^p(x) + \sum_{i=1}^p K(x_i) \big(b_0 m_t^{p-1}((x_j)_{j\neq i}) + b_1 m_t^p((x_j)_{j\neq i}, 0)\big) \\ + \sum_{1\leq i<j\leq p} K(x_i) K(x_j) \big(a_0 m_t^{p-2}((x_k)_{k\neq i,j}) + a_1 m_t^{p-1}((x_k)_{k\neq i,j}, 0) + a_2 m_t^p((x_k)_{k\neq i,j}, 0, 0)\big). \end{aligned}$$

Proof. If we define $M_t(T) = u_t(T-t)$ for $t \leq T$, then :

$$\begin{aligned}
M_t(T) &= \mathbb{E}\left[X_T - \int_t^T K(T-s)b(X_s)ds \middle| \mathcal{F}_t\right], \\
&= \mathbb{E}\left[u_0(T) + \int_0^t K(T-s)b(X_s)ds + \int_0^T K(T-s)\sigma(X_s)dW_s \middle| \mathcal{F}_t\right], \\
&= \int_0^t K(T-s)b(X_s)ds + \mathbb{E}\left[u_0(T) + \int_0^T K(T-s)\sigma(X_s)dW_s \middle| \mathcal{F}_t\right]
\end{aligned}$$

Leading to:

$$M_t(T) = \int_0^t K(T-s)b(X_s)ds + N_t(T),$$

with $N_t(T) = u_0(T) + \int_0^t K(T-s)\sigma(X_s)dW_s$. N is a martingale and M is a semi-martingale, then applying Itô's formula, we get:

$$d\left(\prod_{i=1}^p M_t(T_i)\right) = \sum_{i=1}^p \left[\prod_{j\neq i}^p M_t(T_j)\right] dM_t(T_i) + \sum_{1\leq i<j\leq p} \left[\prod_{k\neq i,j} M_t(T_k)\right] d\langle M(T_i), M(T_j)\rangle_t,$$

with:

$$dM_t(T_i) = K(T_i-t)b(X_t)dt + K(T_i-t)\sigma(X_t)dW_t, \qquad d\langle M(T_i), M(T_j)\rangle_t = K(T_i-t)K(T_j-t)a(X_t)dt,$$

Integrating between 0 and t and taking the expectation, we get:

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^p M_t(T_i)\right] = \prod_{i=1}^p M_0(T_i) &+ \sum_{i=1}^p \int_0^t K(T_i-s)\mathbb{E}\left[\prod_{j\neq i} M_s(T_j)b(X_s)\right] ds, \\
&+ \sum_{1\leq i<j\leq p} \int_0^t K(T_i-s)K(T_j-s)\mathbb{E}\left[\prod_{k\neq i,j} M_s(T_k)a(X_s)\right] ds.
\end{aligned}$$

Taking $T_i = x_i + t$, we remark that $M_t(T_i) = M_t(x_i+t) = u_t(x_i)$. An argument using Itô's formula yields the result.

□

3.3 Moments of Affine Volterra processes

In this section we study the moments of affine Volterra processes which were introduced in Section 1.3. We consider the process X satisfying the stochastic Volterra equation (1.1), that is

$$X_t = X_0 + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \geq 0, \tag{3.9}$$

where $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ is the convolution kernel, $X_0 \in \mathbb{R}^d$, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ continuous and W is a m-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Let $a = \sigma\sigma^\top$ and b be affine maps given by (1.8):

$$\begin{aligned} a(x) &= A_0 + x_1 A_1 + \ldots x_d A_d, \\ b(x) &= b_0 + x_1 b_1 + \ldots x_d b_d, \end{aligned} \tag{3.10}$$

for some $d-$dimensional symmetric matrices A_i and d-dimensional vectors b_i. We set $B = (b_1 \ldots b_d)$.

For completeness we recall Lemma 1.3.2, that gives the first order moment of the affine Volterra processes.

Theorem 3.3.1 (First order moment of affine Volterra processes)**.** *Let X be an affine Volterra process satisfying* (3.9) *with coefficients as in* (3.10)*. Then, its first order moment is given by:*

$$\mathbb{E}[X_t] = \left(\mathbb{I}_d - \int_0^t R_B(u)du\right) X_0 + \left(\int_0^t E_B(u)du\right) b_0,$$

where R_B is the resolvent of $-KB$ and $E_B = K - R_B * K$. (Refer to Section 1.1 for a deep study of stochastic convolutions and resolvents).

Proof. Rewrite (3.9) as

$$X = X_0 + (KB) * X + K * dZ,$$

where $dZ_t = b_0 dt + \sigma(X_t)dW_t$. We convolve the last equation with R_B and subtract for the first one:

$$X - R_B * X = (X_0 - R_B * X_0) + (KB - R_B * (KB)) * X + E_B * dZ,$$

Note that R_B satisfies $KB - R_B * (KB) = -R_B$, then

$$X = (I - R_B * I)\, X_0 + E_B * (b_0 du + \sigma(X)dW)\,.$$

Before taking conditional expectation in the last equation, notice that the local martingale $M_r(t) = \left(\int_0^r E_B(t-s)\sigma\left(X_s\right)dW_s\right)_{0\le r\le t}$, is a martingale. Its quadratic variation satisfies

$$\mathbb{E}\left[|\langle M\rangle_t|\right] \le \int_0^t |E_B(t-u)|^2\, \mathbb{E}[|\sigma\left(X_u\right)|^2]du \le \|E_B\|_{L^2(0,T)} \max_{u\le t} \mathbb{E}[|\sigma\left(X_u\right)|^2],$$

which is finite by (1.7). Finally, taking the conditional expectation gives:

$$\mathbb{E}[X_t|\mathcal{F}_s] = \left(\mathbb{I}_d - \int_0^t R_B(u)du\right) X_0 + \left(\int_0^t E_B(u)du\right) b_0 + \int_0^s E_B(t-u)\sigma(X_u)dW_u, \tag{3.11}$$

for all $s \le t$. In particular, taking $s = 0$ completes the proof. $\square$

Theorem 3.3.2 below shows that the second order moments of affine Volterra processes are polynomials of degree less than or equal to 2 with respect to the initial condition.

Theorem 3.3.2 (Second order moment of affine Volterra processes)**.** *Let X be an affine Volterra process satisfying* (3.9) *with coefficients as in* (3.10)*. Then the following holds:*

$$\begin{aligned}\mathbb{E}[X_t X_t^\top] &= g_B(t)X_0X_0^\top g_B(t)^\top \\ &+ \left(f_B(t)X_0^\top g_B(t)^\top + g_B(t)X_0 f_B(t)^\top + \int_0^t E_B(t-s)(a(g_B(s)X_0) - A_0)E_B(t-s)^\top ds\right) \\ &\qquad + \left(f_B(t)f_B(t)^\top + \int_0^t E_B(t-s)a(f_B(s))E_B(t-s)^\top ds\right), \end{aligned} \tag{3.12}$$

with $f_B(t) = \left(\int_0^t E_B(s)ds\right) b_0$ *and* $g_B(t) = \left(\mathbb{I}_d - \int_0^t R_B(s)ds\right)$*, where* R_B *is the resolvent of* $-KB$ *and* $E_B = K - R_B * K$*.*

Proof. Equation (3.11) shows that $X_t = Y_t(t)$ where $Y(t)$ is the d-dimensional process:

$$Y_s(t) := \mathbb{E}[X_t|\mathcal{F}_s] = f_B(t) + g_B(t)X_0 + \int_0^s E_B(t-u)\sigma(X_u)dW_u, \qquad s \leq t.$$

Integration by parts yields:

$$\begin{aligned} Y_s(t_1)Y_s(t_2)^\top &= Y_0(t_1)Y_0(t_2)^\top + \int_0^s E_B(t_1-u)a(X_u)E_B(t_2-u)^\top du \\ &+ \int_0^s \left(Y_u(t_1)\left(E_B(t_2-u)\sigma(X_u)dW_u\right)^\top + E_B(t_1-u)\sigma(X_u)dW_u Y_u(t_2)^\top\right), \end{aligned} \tag{3.13}$$

for all $s \leq t_1 \wedge t_2$. We conclude from (3.13) that:

$$\begin{aligned} \mathbb{E}[X_t X_t^\top|\mathcal{F}_s] &= Y_0(t)Y_0(t)^\top \\ &+ \int_0^s \left(Y_u(t)\left(E_B(t-u)\sigma(X_u)dW_u\right)^\top + E_B(t-u)\sigma(X_u)dW_u Y_u(t)^\top\right) \\ &+ \int_0^s E_B(t-u)a(X_u)E_B(t-u)^\top du + \int_s^t E_B(t-u)a(Y_s(u))E_B(t-u)^\top du. \end{aligned} \tag{3.14}$$

for all $s \leq t$. In particular, taking $s = 0$ in (3.14), we obtain:

$$\mathbb{E}[X_t X_t^\top] = (f_B(t)+g_B(t)X_0)(f_B(t)+g_B(t)X_0)^\top + \int_0^t E_B(t-u)a(f_B(u)+g_B(u)X_0)E_B(t-u)^\top du,$$

Linearity of the function $a(x) - A_0$ implies (3.12). □

As a consequence of Theorems 3.3.1 and 3.3.2, the moments of first and second order of an affine Volterra process are polynomials in X_0 of degree less than or equal to one and two, respectively. The coefficients of these polynomials are independent of the initial condition X_0. The following theorem shows that this property also holds for higher order moments.

Theorem 3.3.3 (Moments of Affine Volterra Processes)**.** *Let X be an affine Volterra process satisfying (3.9) with coefficients as in (3.10). Then, for any multi-index α with $|\alpha| = n \in \mathbb{N}$, there exist functions $(h^{\alpha}_{\beta}(t))_{|\beta|\leq n}$ independent of X_0 such that:*

$$\mathbb{E}[X^{\alpha}_t] = \sum_{|\beta|\leq n} h^{\alpha}_{\beta}(t) X^{\beta}_0, \quad t \geq 0. \tag{3.15}$$

Proof. Consider the martingales

$$Y_s(t) = \mathbb{E}[X_t|\mathcal{F}_s] = f_B(t) + g_B(t)X_0 + \int_0^s E_B(t-u)\sigma(X_u)dW_u, \qquad s \leq t. \tag{3.16}$$

with $f_B(t) = \left(\int_0^t E_B(u)du\right) b_0$ and $g_B(t) = \left(\mathbb{I}_d - \int_0^t R_B(u)du\right)$.

Given $n \in \mathbb{N}$ and $i_1, \dots, i_n \in \{1, \dots, d\}$, define recursively the processes $F^{i_1,\dots,i_n}_s(t_1, \dots, t_n)$ and $G^{i_1,\dots,i_n}_s(t_1, \dots, t_n)$, for $s \leq \min\{t_1, \dots, t_n\}$, as follows:

$$\begin{aligned}
F^{i_1}_s(t_1) &= 0, \\
G^{i_1}_s(t_1) &= E^{i_1}_B(t_1 - s), \\
F^{i_1,\dots,i_n}_s(t_1,\dots,t_n) &= Y^{i_n}_s(t_n)F^{i_1,\dots,i_{n-1}}_s(t_1,\dots,t_{n-1}) + G^{i_1,\dots,i_{n-1}}_s(t_1,\dots,t_{n-1})a(X_s)E^{i_n}_B(t_n - s)^{\top}, \\
G^{i_1,\dots,i_n}_s(t_1,\dots,t_n) &= (Y^{i_1}_s(t_1)\dots Y^{i_{n-1}}_s(t_{n-1}))E^{i_n}_B(t_n - s) + Y^{i_n}_s(t_n)G^{i_1,\dots,i_{n-1}}_s(t_1,\dots,t_{n-1}),
\end{aligned} \tag{3.17}$$

where E^i_B represents the i-th row of the matrix E_B and Y^i the i-th row of the vector Y. An induction argument using integration by parts yields:

$$\begin{aligned}
Y^{i_1}_s(t_1)\dots Y^{i_n}_s(t_n) = Y^{i_1}_0(t_1)\dots Y^{i_n}_0(t_n) &+ \int_0^s F^{i_1,\dots,i_n}_u(t_1,\dots,t_n)du \\
&+ \int_0^s G^{i_1,\dots,i_n}_u(t_1,\dots,t_n)\sigma(X_u)dW_u,
\end{aligned} \tag{3.18}$$

for all $n \in \mathbb{N}$, $i_1, \dots, i_n \in \{1, \dots, d\}$ and $s \leq \min\{t_1, \dots, t_n\}$.
Taking expectation and applying Fubini's theorem in (3.18) we get:

$$\mathbb{E}[Y^{i_1}_s(t_1)\dots Y^{i_n}_s(t_n)] = Y^{i_1}_0(t_1)\dots Y^{i_n}_0(t_n) + \int_0^s \mathbb{E}[F^{i_1,\dots,i_n}_u(t_1,\dots,t_n)]du, \tag{3.19}$$

for all $n \in \mathbb{N}$, $i_1, \dots, i_n \in \{1, \dots, d\}$ and $s \leq \min\{t_1, \dots, t_n\}$. In particular, taking $s = t_1 \leq \min\{t_2, \dots, t_n\}$ in (3.19) we obtain:

$$\mathbb{E}[X^{i_1}_s Y^{i_2}_s(t_2)\dots Y^{i_n}_s(t_n)] = Y^{i_1}_0(s)Y^{i_2}_0(t_2)\dots Y^{i_n}_0(t_n) + \int_0^s \mathbb{E}[F^{i_1,\dots,i_n}_u(s,\dots,t_n)]du, \tag{3.20}$$

for all $n \in \mathbb{N}$, $i_1, \dots, i_n \in \{1, \dots, d\}$. Another induction argument using the recursion formulas and (3.20), shows that for all $n \in \mathbb{N}$ and $i_1, \dots i_n \in \{1, \dots, d\}$ there exist functions $(h_\beta^{i_1,\dots,i_n}(s, t_2, \dots, t_n))_{|\beta|\le n}$ independent of X_0 such that

$$\mathbb{E}[X_s^{i_1} Y_s^{i_2}(t_2) \dots Y_s^{i_n}(t_n)] = \sum_{|\beta|\le n} h_\beta^{i_1,\dots,i_n}(s, t_2, \dots, t_n) X_0^\beta, \tag{3.21}$$

for all $s \le \min\{t_2, \dots, t_n\}$. The expression (3.15) follows after taking $s = t2 = \dots = t_n = t$ in (3.21). □

3.3.1 Examples of second and third order moments of X_t in dimension $d = 1$

In this section, we show the explicit form of the polynomial functions on the initial state X_0, which describe the second and third moments of affine Volterra process in Theorem 3.3.3. Also, in the classical case we recover the system of linear differential equations that solve the coefficients that accompany X_0. In all the examples, we fix the dimension $d = 1$.

Example 3.3.4 (Second order moment of X_t)**.** *We have:*

$$\mathbb{E}[X_t^2] = \bar{\alpha}(t) + \tilde{\beta}(t) X_0 + \bar{\gamma}(t) X_0^2, \quad t \ge 0,$$

with:

$$\begin{aligned}
\bar{\alpha}(t) =& f_B^2(t) + \int_0^t E_B^2(u)(A_0 + A_1 f_B(t-u))du. \\
\bar{\beta}(t) =& 2 f_B(t) g_B(t) + A_1 \int_0^t E_B^2(u) g_B(t-u) du. \\
\bar{\gamma}(t) =& g_B^2(t).
\end{aligned}$$

Recall that $f_B'(t) = E_B(t) b_0$, $g_B'(t) = E_B(t) b_1$ with initial conditions $g_B(0) = 0$ and $f_B(0) = 1$. Differentiation with respect to t, yields:

$$\bar{\alpha}'(t) = b_0 \left(2 f_B(t) E_B(t) + A_1 \int_0^t E_B^2(u) E_B(t-u) du \right) + A_0 E_B^2(t). \tag{3.22}$$

$$\bar{\beta}'(t) = b_1 \left(2 f_B(t) E_B(t) + A_1 \int_0^t E_B^2(u) E_B(t-u) du \right) + 2 b_0 g_B(t) E_B(t) + A_1 E_B^2(t). \tag{3.23}$$

$$\bar{\gamma}'(t) = 2 b_1 g_B(t) E_B(t). \tag{3.24}$$

Classical case $K \equiv 1$

In the classical affine diffusion case, when $K \equiv 1$ and $g_B = E_B$, equations (3.22), (3.23) and (3.24) can be written as a linear differential system in $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$,

$$\begin{aligned}
\bar{\alpha}'(t) =& b_0\bar{\beta}(t) + A_0\bar{\gamma}(t).\\
\bar{\beta}'(t) =& b_1\bar{\beta}(t) + (2b_0 + A_1)\bar{\gamma}(t).\\
\bar{\gamma}'(t) =& 2b_1\bar{\gamma}(t).
\end{aligned}$$

The matrix of coefficients for the moments of order up to two has the following form

$$\Gamma(t) = \begin{pmatrix} 0 & f_B(t) & \bar{\alpha}(t) \\ 0 & g_B(t) & \bar{\beta}(t) \\ 0 & 0 & \bar{\gamma}(t) \end{pmatrix}$$

and solves the linear differential equation:

$$\Gamma'(t) = G\Gamma(t), \qquad \Gamma(0) = \mathbb{I}_3,$$

where G is the matrix of the infinitesimal generator of X restricted to Pol_2 and is given by:

$$\Gamma = \begin{pmatrix} 0 & b_0 & A_0 \\ 0 & b_1 & 2b_0 + A_1 \\ 0 & 0 & 2b_1 \end{pmatrix}$$

Example 3.3.5 (Third order moment of X_t)**.** *We have:*

$$\mathbb{E}[X_t^3] = \tilde{\alpha}(t) + \tilde{\beta}(t)X_0 + \tilde{\gamma}(t)X_0^2 + \tilde{\zeta}(t)X_0^3, \quad t \geq 0,$$

with:

$$\begin{aligned}
\tilde{\alpha}(t) =& f_B^3(t) + 3A_0 f_B(t)\int_0^t E_B^2(t-u)du + 3A_1 f_B(t)\int_0^t E_B^2(t-u)f_B(u)du\\
&+ 3A_1\int_0^t E_B^2(t-u)\int_0^u E_B(u-r)E_B(t-r)(A_0 + A_1 f_B(r))drdu.\\
\tilde{\beta}(t) =& 3f_B^2(t)g_B(t) + 3A_0 g_B(t)\int_0^t E_B^2(t-u)du + 3A_1 f_B(t)\int_0^t E_B^2(t-u)g_B(u)du\\
&+ 3A_1 g_B(t)\int_0^t E_B^2(t-u)f_B(u)du + 3A_1\int_0^t E_B^2(t-u)\int_0^u E_B(u-r)E_B(t-r)A_1 g_B(r)drdu.\\
\tilde{\gamma}(t) =& 3f_B(t)g_B^2(t) + 3A_1 g_B(t)\int_0^t E_B^2(t-u)g_B(u)du.\\
\tilde{\zeta}(t) =& g_B^3(t).
\end{aligned}$$

Proof. Following the general formula (3.35), we have:

$$\mathbb{E}[X_t^3] = \big(Y_0(t)\big)^3 + 3\int_0^t E_B^2(t-u)\mathbb{E}[Y_u(t)a(X_u)]du. \tag{3.25}$$

Taking the expectation term in the right, we obtain:

$$\begin{aligned}\mathbb{E}[Y_u(t)a(X_u)] = {} & A_0\mathbb{E}[Y_u(t)] + A_1\mathbb{E}[Y_u(t)X_u] \\ = {} & A_0Y_0(t) + A_1\mathbb{E}[Y_u(t)Y_u(u)] \\ = {} & A_0Y_0(t) + A_1\left(Y_0(u)Y_0(t) + \int_0^u E_B(t-r)E_B(u-r)a\left(f_B(r)+g_B(r)x_0\right)dr\right).\end{aligned} \tag{3.26}$$

Plugging (3.26) in (3.25), yields:

$$\begin{aligned}\mathbb{E}[X_t^3] = {} & \big(Y_0(t)\big)^3 + 3\int_0^t E_B^2(t-u)A_0Y_0(t)du + 3A_1\int_0^t E_B^2(t-u)Y_0(t)Y_0(u)du \\ & + 3A_1\int_0^t E_B^2(t-u)\left(\int_0^u E_B(u-r)E_B(t-r)a\left(f_B(r)+g_B(r)x_0\right)dr\right)du.\end{aligned} \tag{3.27}$$

If we remember that $Y_0(t) = f_B(t) + g_B(t)X_0$, then

$$\begin{aligned}\mathbb{E}[X_t^3] = {} & f_B^3(t)+3f_B^2(t)g_B(t)x_0+3f_B(t)g_B^2(t)x_0^2+g_B^3(t)x_0^3+3A_0(f_B(t)+g_B(t)x_0)\int_0^t E_B^2(t-u)du \\ & + 3A_1(f_B(t)+g_B(t)x_0)\int_0^t E_B^2(t-u)(f_B(u)+g_B(u)x_0)du \\ & + 3A_1\int_0^t E_B^2(t-u)\int_0^u E_B(u-r)E_B(t-r)\left(A_0+A_1(f_B(r)+g_B(r)x_0)\right)drdu.\end{aligned} \tag{3.28}$$

Regrouping the terms in (3.28) according to the powers of x_0 we finish the proof. □

Differentiation of $\tilde{\alpha}(t)$, $\tilde{\beta}(t)$, $\tilde{\gamma}(t)$ and $\tilde{\zeta}(t)$ with respect to t, remembering that $f_B'(t) = E_B(t)b_0$, $g_B'(t) = E_B(t)b_1$ with initial conditions $g_B(0) = 0$ and $g_B(0) = 1$, yields:

$$\begin{aligned}\tilde{\alpha}'(t) = {} & 3b_0f_B^2(t)E_B(t) + 3b_0A_0E_B(t)\int_0^t E_B^2(t-u)du + 3A_0f_B(t)E_B^2(t) \\ & + 3b_0A_1E_B(t)\int_0^t E_B^2(t-u)f_B(u)du + 3A_1b_0f_B(t)\int_0^t E_B^2(t-u)E_B(u)du \\ & + 3b_0A_1\int_0^t E_B^2(t-u)\int_0^u E_B(t-r)E_B(u-r)A_1E_B(r)drdu + 3A_0A_1E_B(t)\int_0^t E_B^2(t-u)E_B(u)du.\end{aligned} \tag{3.29}$$

$$\begin{aligned}\tilde{\beta}'(t) = & \ 6b_0 f_B(t)E_B(t)g_B(t) + 3b_1 f_B^2(t)E_B(t) + 3b_1 A_0 E_B(t)\int_0^t E_B^2(t-u)du + 3A_0 g_B(t)E_B^2(t) \\ & + 3b_0 A_1 E_B(t)\int_0^t E_B^2(t-u)g_B(u)du + 3A_1 b_1 f_B(t)\int_0^t E_B^2(t-u)E_B(u)du + 3A_1 f_B(t)E_B^2(t) \\ & + 3b_1 A_1 E_B(t)\int_0^t E_B^2(t-u)f_B(u)du + 3A_1 b_0 g_B(t)\int_0^t E_B^2(t-u)E_B(u)du \\ & + 3A_1 b_1 \int_0^t E_B^2(t-u)\int_0^u E_B(u-r)E_B(t-r)A_1 E_B(r)drdu + 3A_1^2 E_B(t)\int_0^t E_B^2(t-u)E_B(u)du.\end{aligned} \tag{3.30}$$

$$\begin{aligned}\tilde{\gamma}'(t) = & 3b_0 E_B(t)g_B^2(t) + 6b_1 f_B(t)g_B(t)E_B(t) + 3A_1 b_1 E_B(t)\int_0^t E_B^2(t-u)g_B(u)du \\ & + 3A_1 b_1 g_B(t)\int_0^t E_B(t-u)E_B^2(u)du + 3A_1 g_B(t)E_B^2(t).\end{aligned} \tag{3.31}$$

$$\tilde{\zeta}'(t) = 3b_1 g_B^2(t)E_B(t). \tag{3.32}$$

Classical case $K \equiv 1$

In the classical affine diffusion case when $K \equiv 1$ and $g_B = E_B$, equations (3.29), (3.30), (3.31) and (3.3.1) can be written as the following linear differential system:

$$\begin{aligned}\tilde{\alpha}'(t) = & b_0\tilde{\beta}(t) + A_0\tilde{\gamma}(t). \\ \tilde{\beta}'(t) = & b_1\tilde{\beta}(t) + (2b_0 + A_1)\tilde{\gamma}(t) + 3A_0\tilde{\zeta}(t). \\ \tilde{\gamma}'(t) = & 3(b_0 + A_1)\tilde{\zeta}(t) + 2b_1\tilde{\gamma}(t). \\ \tilde{\zeta}'(t) = & 3b_1\tilde{\zeta}(t).\end{aligned}$$

The matrix of coefficients for the moments of order less than or equal to 3, solves the following linear system:

$$\Gamma'(t) = G\Gamma(t), \qquad \Gamma(0) = \mathbb{I}_4,$$

$$\begin{pmatrix} 0 & f_B'(t) & \alpha'(t) & \tilde{\alpha}'(t) \\ 0 & g_B'(t) & \beta'(t) & \tilde{\beta}'(t) \\ 0 & 0 & \gamma'(t) & \tilde{\gamma}'(t) \\ 0 & 0 & 0 & \tilde{\zeta}'(t) \end{pmatrix} = \begin{pmatrix} 1 & b_0 & A_0 & 0 \\ 0 & b_1 & 2b_0 + A_1 & 3A_0 \\ 0 & 0 & 2b_1 & 3(A_1 + b_0) \\ 0 & 0 & 0 & 3b_1 \end{pmatrix} \begin{pmatrix} 0 & f_B(t) & \alpha(t) & \tilde{\alpha}(t) \\ 0 & g_B(t) & \beta(t) & \tilde{\beta}(t) \\ 0 & 0 & \gamma(t) & \tilde{\gamma}(t) \\ 0 & 0 & 0 & \tilde{\zeta}(t) \end{pmatrix}$$

The matrix G is the infinitesimal generator of X restricted to Pol_3.

3.3.2 Conditional moments in dimension $d = 1$

In this part, we prove a recursive formula to compute the conditional moments of affine Volterra process.

Theorem 3.3.6 (General formula for high-order conditional moments)**.** *Let X be an affine Volterra process* (3.9) *with $d = 1$. Then, for $p \geq 2$:*

$$\mathbb{E}[X_t^p|\mathcal{F}_s] = \mathbb{E}[X_t|\mathcal{F}_s]^p + \frac{p(p-1)}{2}\int_s^t E_B^2(t-u)\mathbb{E}[(Y_u(t))^{p-2}a(X_u)|\mathcal{F}_s]du \qquad s \leq t,$$

where $Y_s(t) = \mathbb{E}[X_t|\mathcal{F}_s]$.

Proof. Consider the martingale:

$$Y_s(t) = Y_0(t) + \int_0^s E_B(t-u)\sigma(X_u)dW_u, \quad s \leq t. \tag{3.33}$$

with $Y_0(t) = f_B(t) + g_B(t)X_0$. Applying Ito's formula to the precedent equation we get:

$$d(Y_s(t))^p \;\; = \;\; p\,(Y_s(t))^{p-1}\,E_B(t-s)\sigma(X_s)dW_s + \frac{1}{2}p(p-1)\,(Y_s(t))^{p-2}\,E_B^2(t-s)a(X_s)ds.$$

In integral form:

$$(Y_s(t))^p = (Y_0(t))^p + \int_0^s \frac{1}{2}p(p-1)(Y_u(t))^{p-2}E_B^2(t-u)a(X_u)du + \int_0^s p(Y_u(t))^{p-1}E_B(t-u)\sigma(X_u)dW_u. \tag{3.34}$$

Taking $s = t$ in (3.34) leads to:

$$X_t^p = (Y_0(t))^p + \int_0^t \frac{1}{2}p(p-1)(Y_u(t))^{p-2}E_B^2(t-u)a(X_u)du + \int_0^t p(Y_u(t))^{p-1}E_B(t-u)\sigma(X_u)dW_u.$$

Noticing that the second integral is a martingale and applying conditional expectation to the precedent equation, we get:

$$\mathbb{E}[X_t^p|\mathcal{F}_s] = \mathbb{E}[X_t|\mathcal{F}_s]^p + \frac{p(p-1)}{2}\int_s^t E_B^2(t-u)\mathbb{E}[(Y_u(t))^{p-2}a(X_u)|\mathcal{F}_s]du, \qquad s \leq t, \tag{3.35}$$

where $(Y_s(t))^p = \mathbb{E}[X_t|\mathcal{F}_s]^p$.

∎

3.3.3 Conditional moments using the characteristic function with $d = 1$

When $d = 1$, the conditional characteristic function of an affine process X in Theorem 1.3.1 is given by:

$$\mathcal{L}(u) = \mathbb{E}[\exp(uX_t)\,|\mathcal{F}_s] = \exp\left(u\mathbb{E}[X_t|\mathcal{F}_s] + \frac{1}{2}\int_s^t a(\mathbb{E}[X_r|\mathcal{F}_s])\psi^2(t-r,u)dr\right), \tag{3.36}$$

for $s \leq t$, where ψ satisfies the following Volterra Riccati equation:

$$\psi(t-s,u) = uE_B(t-s) + \frac{A_1}{2}\big(\psi^2(.,u) * E_B(.)\big)(t-s), \tag{3.37}$$

with initial conditions $\psi(t,0) = 0$ for all $t \in [0,T]$.

The conditional characteristic function (3.36) is continuously differentiable and the conditional moments of the process X_t are given by :

$$\mathbb{E}[X_t^p|\mathcal{F}_s] = \mathcal{L}^{(p)}(0), \qquad s \leq t, \tag{3.38}$$

where $\mathcal{L}^{(p)}$ represents the p-th derivative of the conditional characteristic function.

We are going to express the characteristic function (3.36) in a compact way. In order to do that, we define the process $g_s(t) = \mathbb{E}[X_{s+t}|\mathcal{F}_s]$. Then,

$$\mathcal{L}(u) = exp\left(ug_s(t-s) + \frac{1}{2}\big(a\big(g_s(.)\big) * \psi^2(.,u)\big)(t-s)\right). \tag{3.39}$$

Let $\mathcal{N}(u) = ug_s(t-s) + \frac{1}{2}\left(a\left(g_s(.)\right) * \psi^2(.,u)\right)(t-s)$.

In the following $\mathcal{L}^{(p)}$ and $\mathcal{N}^{(p)}$ represent the p-th derivatives of $\mathcal{L}$ and $\mathcal{N}$, respectively. To ease notation let $\psi^{(p)}(.,u)$ represent $\frac{\partial^p}{\partial u^p}\psi(.,u)$. Then, thanks to the Leibniz formula, for $p \geq 2$ we get:

$$\mathcal{L}^{(p)}(u) = \sum_{k=0}^{p-1}\binom{p-1}{k}\mathcal{L}^{(k)}(u)\mathcal{N}^{(p-k)}(u), \tag{3.40}$$

$$\mathcal{N}^{(p)}(u) = \frac{1}{2}\left(a\big(g_s(.)\big) * \sum_{k=0}^{p}\binom{p}{k}\psi^{(k)}(.,u)\psi^{(p-k)}(.,u)\right)(t-s), \tag{3.41}$$

$$\psi^{(p)}(t-s,u) = \frac{A_1}{2}\left(\sum_{k=0}^{p}\binom{p}{k}\psi^{(k)}(.,u)\psi^{(p-k)}(.,u) * E_B(.)\right)(t-s), \tag{3.42}$$

with initial conditions:

$$\mathcal{L}(0) = 1, \qquad \mathcal{L}^{(1)}(u) = \mathcal{L}(u)\mathcal{N}^{(1)}(u), \tag{3.43}$$

$$\mathcal{N}(0) = 0, \qquad \mathcal{N}^{(1)}(u) = g_s(t-s) + \frac{1}{2}\left(a\big(g_s(.)\big) * \frac{\partial}{\partial_u}\psi^2(.,u)\right)(t-s), \tag{3.44}$$

$$\psi(t-s,0) = 0, \qquad \psi^{(1)}(t-s,u) = E_B(t-s) + \frac{A_1}{2}\left(\frac{\partial}{\partial_u}\psi^2(.,u) * E_B(.)\right)(t-s) \tag{3.45}$$

The p-th conditional moment of the process X_t is given evaluating $\mathcal{L}^{(p)}(0)$. For, $p \geq 2$,

$$\mathcal{L}^{(p)}(0) = \sum_{k=0}^{p-1} \binom{p-1}{k} \mathcal{L}^{(k)}(0)\mathcal{N}^{(p-k)}(0), \tag{3.46}$$

with,

$$\mathcal{N}^{(p)}(0) = \frac{1}{2}\left(a\big(g_s(.)\big) * \sum_{k=1}^{p-1} \binom{p}{k} \psi^{(k)}(.,0)\psi^{(p-k)}(.,0) \right)(t-s),$$
$$\psi^{(p)}(t-s,0) = \frac{A_1}{2}\left(\sum_{k=1}^{p-1} \binom{p}{k} \psi^{(k)}(.,0)\psi^{(p-k)}(.,0) * E_B(.) \right)(t-s),$$

and initial conditions:

$$\mathcal{L}(0) = 1, \qquad \mathcal{L}^{(1)}(0) = g_s(t-s), \tag{3.47}$$
$$\mathcal{N}(0) = 0, \qquad \mathcal{N}^{(1)}(0) = g_s(t-s), \tag{3.48}$$
$$\psi(t-s,0) = 0, \qquad \psi^{(1)}(t-s,0) = E_B(t-s). \tag{3.49}$$

Examples

- First order conditional moment:

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathcal{L}^{(1)}(0), \quad s \leq t.$$

- Second order conditional moment:

$$\mathbb{E}[X_t^2|\mathcal{F}_s] = \mathcal{L}^{(2)}(0) = (\mathbb{E}[X_t|\mathcal{F}_s])^2 + \int_s^t a(\mathbb{E}[X_r|\mathcal{F}_s])E_B^2(t-r)dr, \quad s \leq t.$$

- Third order conditional moment:

$$\begin{aligned}\mathbb{E}[X_t^3|\mathcal{F}_s] = \mathcal{L}^{(3)}(0) &= \mathbb{E}[X_t^2|\mathcal{F}_s]\mathbb{E}[X_t|\mathcal{F}_s] + 2\mathbb{E}[X_t|\mathcal{F}_s]\int_s^t a(\mathbb{E}[X_r|\mathcal{F}_s])E_B^2(t-r)dr \\ &+ 3A_1 \int_s^t a(\mathbb{E}[X_r|\mathcal{F}_s])E_B(t-r)(E_B^2 * E_B)(t-r)dr, \quad s \leq t.\end{aligned}$$

3.4 Numerical Implementation

3.4.1 Generator matrix for moments using factor-approximation

We will use the factor-approximation studied in Section 1.4.2 to approach a process satisfying the stochastic Volterra equation (3.9). This factor approximation enjoying of Markovian and semi-martingale properties will allow us to compute the moments through evaluation of an exponential matrix as studied in Section 3.1.

We recall that a completely monotone kernel (as the fractional kernel) can be seen as the Laplace transform of a positive measure μ, that is

$$K(t) = \int_{\mathbb{R}_+} e^{-\gamma t} \mu(d\gamma) \quad t \geq 0.$$

We can then, approximate K by a sequence of smoothed kernels $(K^n)_{n\in\mathbb{N}}$ of the form:

$$K^n(t) = \sum_{i=1}^{n} c_i e^{-\gamma_i t} \text{ for }, \quad t \geq 0. \tag{3.50}$$

This leads to the following approximation X^n of the process X satisfying the stochastic Volterra equation (3.9). For $n \in \mathbb{N}$ and $t \geq 0$, we have:

$$X_t^n = X_0 + \sum_{i=1}^{n} c_i Y_t^{n,i}, \tag{3.51}$$

where for any $i \in \{1, ..., n\}$, Y^n is the solution of the following SDE:

$$dY_t^{n,i} \;=\; \big(-\gamma_i Y_t^{n,i} + b(X_t^n)\big)dt + \sigma(X_t^n)dW_t^n, \qquad Y_0^{n,i} = 0 \text{ for all i} \in \{1, 2, \ldots, n\} \tag{3.52}$$

The convergence of the moments of X_t^n in (3.51) to the moments of X_t is ensured by the following result.

Theorem 3.4.1 (Convergence of moments)**.** *Suppose that X^n is a sequence of random variables in $\mathbb{R}^d$ that converges weakly to X and for every $k \geq 2$*

$$\sup_{n\geq 1} \mathbb{E}[|X^n|^k] < \infty. \tag{3.53}$$

Then,

$$\mathbb{E}[P(X^n)] \to \mathbb{E}[P(X)], \quad \textit{for every polynomial } P \textit{ on } \mathbb{R}^d.$$

Proof. Applying the continuous mapping theorem we get that $P(X^n)$ converges weakly to $P(X)$. Thanks to (3.53), De la Vallé Poussin criterion implies that $P(X^n)$ is uniformly integrable. The conclusion follows from [Billingsley, 1999, Theorem 3.5].

□

We remark that Theorem 1.4.4 gives the conditions to satisfy the assumptions in Theorem (3.4.1).

Now, we fix d=1 and recall the form of the function b and $a = \sigma^2$ in (3.10):

$$b(x) = b_0 + b_1 x \quad \text{and} \quad a(x) = A_0 + A_1 x + A_2 x^2,$$

with b_0, b_1, A_0, A_1 and $A_2 \in \mathbb{R}$. Developing (3.52), we get:

$$dY_t^{n,i} = \left(-\gamma_i Y_t^{n,i} + b_0 + b_1 X_0 + b_1 \sum_{j=1}^{n} c_j Y_t^{n,j}\right) dt + \sigma \left(X_0 + \sum_{j=1}^{n} c_j Y_t^{n,j}\right) dW_t. \tag{3.54}$$

for all i $\in \{1, 2, \ldots, n\}$, which leads to the following representation:

$$dY_t^n = \left(-\text{diag}(\gamma) Y_t^n + (b_0 + b_1 X_0 + b_1 c^\top Y_t^n)\mathbf{1}\right) dt + \sigma \left(X_0 + c^\top Y_t^n\right) \mathbf{1} dW_t, \tag{3.55}$$

with initial value $Y_0^n = 0$, where diag(γ) is the diagonal matrix containing the parameters $(\gamma_1, \ldots, \gamma_n)$ in main diagonal, c is the column vector which entries correspond to the weights c_i, for $i \in \{1, 2, \ldots, n\}$ and $\mathbf{1}$ is a column vector with all its entries equal to 1.

We remark that $Y^n = (Y^{(n,i)})_{i=1}^n$ is polynomial in the classical sense. We remember that the affine framework is a particular case of polynomial diffusions studied in Section 3.1. Based on the classical polynomial diffusion approach where we can compute the moments in an explicit form using an exponential matrix, we give a formula to compute the moments of Y^n.

Let

$$A = \begin{pmatrix} c_1 & c_2 & c_3 & \ldots & c_N \\ 1 & -1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

and define $\hat{Y}^n = AY^n$. The process $\hat{Y}^n$ satisfies the following stochastic differential equation:

$$d\hat{Y}_t^n = \left(-A\text{diag}(\gamma)A^{-1}\hat{Y}_t^n + \left(b_0 + b_1 X_0 + b_1 \hat{Y}_t^{n,1}\right) A\mathbf{1}\right) dt + \sigma \left(X_0 + \hat{Y}_t^{n,1}\right) A\mathbf{1} dW_t, \tag{3.56}$$

Observe that $\hat{Y}^{n,1}$ corresponds to the first coordinate of $\hat{Y}^n$ and is equal to $\sum_{i=1}^n c_i Y^{n,i}$. If we let:

$$\begin{aligned} V &= (b_0 + b_1 X_0) A\mathbf{1}. \\ \Gamma &= -A\mathrm{diag}(\gamma)A^{-1} + b_1 A\mathbf{1}e_1^\top. \\ \Theta &= \sigma(X_0 + \hat{Y}^{n,1})A\mathbf{1}\mathbf{1}^\top A^\top, \end{aligned}$$

where e_1 is a vector of size n with first entry equal to 1 and the others equal to 0. Now, we can rewrite (3.56) as

$$d\hat{Y}_t^n = (V + \Gamma\hat{Y}_t^n)dt + \Theta_t dW_t.$$

Then, the partial differential operator applied to a polynomial f will be given by:

$$\mathcal{G}^{\hat{Y}^n} f(y) = \nabla f(y)^\top (V + \Gamma y) + \frac{1}{2} a\,(x_0 + y) \left(\sum_{i=1}^n c_i \right)^2 \partial^2_{y^1 y^1} f(y).$$

And now, we will be able to compute the moments of the process X. Let the function $f(y) = y^\alpha$, where α is a multi-index of the form $y^\alpha = (y^1)^{\alpha_1}(y^2)^{\alpha_2}\ldots(y^n)^{\alpha_n}$ with $\alpha_1 + \alpha_2 + \ldots \alpha_n \le m$. It means that with the generator matrix we can determinate the moments of X^n of degree at most m. Remark that the moment of order p of $X^n - X_0$ is given by the corresponding row $(\hat{Y}^{n,1})^p$.

3.4.2 Adam's method

We follow the ideas behind the Adam's method [Li and Tao, 2009], which have bee proved very useful to solve Riccati fractional equations, to find the moments of a process X satisfying (3.4). We consider the fractional kernel $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ with $\alpha \in (1/2, 1]$ and apply Adam's method to solve (3.8). We develop the procedure for the first order moment to understand the complexity of this method to solve even the first order moment which proves to be cumbersome for higher order of moments.

First order moment

Let $d = 1$. We are going to use Adam's method to find the first order moment of a process X satisfying (3.4) with the fractional kernel $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$, $\alpha \in (1/2, 1]$. When $p = 1$, equation (3.8) is given by:

$$\begin{aligned} m(t,x) &= u_0(x+t) + \int_0^t K(t+x-s)[b_0 + b_1 m(s,0)]ds, \\ &= u_0(x+t) + b_0 \int_0^t \frac{(t+x-s)^{\alpha-1}}{\Gamma(\alpha)} ds + b_1 \int_0^t \frac{(t+x-s)^{\alpha-1}}{\Gamma(\alpha)} m(s,0)ds, \\ &= \underbrace{u_0(x+t) + \frac{b_0}{\alpha\Gamma(\alpha)}(-x^\alpha + (t+x)^\alpha)}_{\mathcal{L}(x,t)} + b_1 \int_0^t \frac{(t+x-s)^{\alpha-1}}{\Gamma(\alpha)} m(s,0)ds. \end{aligned} \tag{3.57}$$

Following the Adam's method, we do a linear interpolation:

$$\hat{m}(t,0) = \frac{t_{j+1}-t}{t_{j+1}-t_j}\hat{m}(t_j,0) + \frac{t-t_j}{t_{j+1}-t_j}\hat{m}(t_{j+1},0) \quad 1_{t_j \le t \le t_{j+1}},$$

Then,

$$\begin{aligned} \hat{m}(t_{k+1},x) = \mathcal{L}(x,t_{k+1}) + \frac{b_1}{\Gamma(\alpha)} \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} & (t_{k+1}+x-s)^{\alpha-1} \frac{t_{j+1}-s}{t_{j+1}-t_j}\hat{m}(t_j,0) \\ & + (t_{k+1}+x-s)^{\alpha-1} \frac{s-t_j}{t_{j+1}-t_j}\hat{m}(t_{j+1},0)ds. \end{aligned}$$

Developing the last expression, Adam's method gives a solution to (3.57) in the following implicit schema:

$$\begin{aligned} &\hat{m}(t_{k+1},x) = \mathcal{L}(x,t_{k+1}) + \\ &\frac{b_1}{\Gamma(\alpha)} \sum_{j=0}^{k} \left[\frac{(t_{j+1}-t_j)(t_{k+1}+x-t_j)^\alpha}{\Delta\alpha} + \frac{(t_{k+1}+x-t_{j+1})^{\alpha+1}}{\Delta\alpha(\alpha+1)} - \frac{(t_{k+1}+x-t_j)^{\alpha+1}}{\Delta\alpha(\alpha+1)} \right] \hat{m}(t_j,0) \\ &+ \left[\frac{-(t_{j+1}-t_j)(t_{k+1}+x-t_{j+1})^\alpha}{\Delta\alpha} - \frac{(t_{k+1}+x-t_{j+1})^{\alpha+1}}{\Delta\alpha(\alpha+1)} + \frac{(t_{k+1}+x-t_j)^{\alpha+1}}{\Delta\alpha(\alpha+1)} \right] \hat{m}(t_{j+1},0). \end{aligned} \tag{3.58}$$

When $j=k$ and $x=0$ we have $\hat{m}(t_{k+1},0)$ in both sides of the equation, so that $\hat{m}(t_{k+1},0)$ in the right side will be estimated by a predictor given by:

$$\hat{m}^P(t_{k+1},0) = \mathcal{L}(0,t_{k+1}) + \frac{b_1}{\Gamma(\alpha)} \int_0^{t_{k+1}} \frac{(t_{k+1}-s)^{\alpha-1}}{\Gamma(\alpha)} \tilde{m}(s,0)ds,$$

with $\tilde{m}(t,0) = \sum_{j=0}^{k} \hat{m}(t,0) 1_{t_j \le t \le t_{j+1}}$, so that:

$$\hat{m}^P(t_{k+1},0) = \mathcal{L}(0,t_{k+1}) + \frac{b_1}{\Gamma(\alpha)} \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} (t_{k+1}-s)^{\alpha-1} \hat{m}(t_j,0)ds.$$

Finally, developing the last expression, the predictor is given by:

$$\hat{m}^P(t_{k+1},0) = g_0(t_{k+1}) + \frac{b_0}{\alpha\Gamma(\alpha)}(t_{k+1})^\alpha + \frac{b_1}{\Gamma(\alpha)}\sum_{j=0}^{k}\left[\frac{(t_{k+1}-t_j)^\alpha - (t_{k+1}-t_{j+1})^\alpha}{\alpha}\right]\hat{m}(t_j,0)ds. \tag{3.59}$$

Equations (3.58) and (3.59) give the Adam's scheme to find the first order moment of the process X when $x = 0$ with initial condition $\hat{m}(t_0, 0) = u_0(0)$.

3.4.3 Numerical Results

We compute the moments of a Volterra CIR process V satisfying:

$$V_t = V_0 - \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\left(\lambda(\bar{\nu} - V_s)\,\mathrm{d}s + \eta\sqrt{V_s}\,\mathrm{d}W_s\right) \quad t \leq T, \tag{3.60}$$

with parameters:

$$\lambda = 2 \quad \bar{\nu} = 4 \quad \eta = 0.3 \quad V_0 = 4 \quad T = 1.5 \quad \alpha = 0.6$$

The Adam's method approach proved to be time consuming and do not give a good approximation for the moments compared to the matrix generator. Therefore, we do not include the results here. We present the computation results for the first, second and third order moments of V in Tables (3.1), (3.2) and (3.3), respectively. We show the results using the matrix generator approach and Monte Carlo simulation. The first column represents the number of factors used in the multi-factor approximation of the process V in order to apply our generator matrix approach. The second column gives the value of the respective moment. The third column has the number of Monte Carlo simulations and the last column shows the respectively computed moment.

N. of factors n	Matrix Generator approach	N.of MC simulations	MC
10	4	10^5	3.9992
20	4	10^5	4.0042

Table 3.1: Monte Carlo simulation vs Matrix Generator approach first order moment computation.

N. of factors n	Matrix Generator approach	N.of MC simulations	MC
10	16.1649	10^5	16.2322
20	16.2870	10^5	16.2732

Table 3.2: Monte Carlo vs Matrix Generator approach second order moment computation.

N. of factors n	Matrix Generator approach	N.of MC simulations	MC
10	65.9922	10^5	66.8456
20	67.4820	10^5	67.1013

Table 3.3: Monte Carlo vs Matrix Generator approach third order moment computation.

Chapter 4

Estimation of the drift for a Volterra Ornstein-Uhlenbeck process

In this chapter, we focus on the statistical issue of parameter estimation in the Volterra version of the Ornstein-Uhlenbeck process. The Volterra Ornstein-Uhlenbeck process is the unique Gaussian process X satisfying the stochastic convolution equation

$$X_t = \int_0^t K(t-s)(\theta X_s ds + \sigma dW_s), \quad t \geq 0, \tag{4.1}$$

where $\theta \in \mathbb{R}$, $\sigma^2 > 0$, W is a standard Brownian motion in a complete probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and K is the fractional kernel defined on $(0, +\infty)$ by

$$K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}, \quad \alpha \in \left(\frac{1}{2}, 1\right]. \tag{4.2}$$

The constant drift and the variance, θ and σ^2, are unknown parameters. We investigate their estimation based on the observation of the process X in continuous and discrete time. To the best of our knowledge, this is the first work addressing parameter estimation in the Volterra Ornstein-Uhlenbeck process.

The variance parameter σ^2 will be computed in a relatively easy way. We therefore concentrate on the task of the estimation of the drift parameter θ. The case when $\alpha = 1$ in the fractional kernel (4.2), corresponds to the well-documented drift parameter estimation for the Ornstein-Uhlenbeck process [Kutoyants, 2004, Liptser and Shiryaev, 2001, Basawa and Scott, 1983]. There has been special attention to the study of the fractional version of the Ornstein-Uhlenbeck, which is a continuous Gaussian process satisfying the stochastic differential equation

$$dX_t = \theta X_t dt + dW_t^H \tag{4.3}$$

where W^H is the fractional Brownian motion in (22) with Hurst parameter $H \in (0,1)$. We remark that this process is similar in form to ours. In fact the Riemann-Liouville version of the fractional Brownian motion is given by

$$\int_0^t K(t-s)dW_s$$

where K the fractional kernel (4.2). The relation between α and H is given by $\alpha = H + 1/2$. The main difference between the fractional and the Volterra Ornstein-Uhlenbeck processes is the appearance of the fractional kernel next to the drift parameter.

The problem of estimation of the drift parameter in the fractional Ornstein-Uhlenbeck process (4.3) has its difficulties. This is due mainly to the fractional Brownian motion is neither a semimartingale nor a Markovian process for $H \neq 1/2$. This problem has been addressed in many works. The most popular estimators are the maximum likelihood and the least squares estimators. When it is assumed that one observes the whole trajectory of X in continuous time, strong consistency for both estimators has been proved. In Kleptsyna and Le Breton [2002], they consider the maximum likelihood estimator of $\theta \in \mathbb{R}$ when $H \geq 1/2$. In Hu and Nualart [2010], they study the least squares estimator of $\theta < 0$ when $H \geq 1/2$ (a complementary work for $H \in (0,1)$ is done in Hu et al. [2019]). The least squares estimator for $\theta > 0$ is analyzed in Belfadli et al. [2011] when $H > 1/2$. This last study is generalized for $H \in (0,1)$ in El Machkouri et al. [2016]. In Tudor and Viens [2007], they prove the strong consistency of the maximum likelihood estimator when the process is observed in continuous time and that a version of the this estimator is still strong consistency using only discrete observations for $\theta \in \mathbb{R}$ with $H \in (0,1)$.

A discretized version of the least squares estimator is considered in Cénac and Es-Sebaiy [2015], Es-Sebaiy and Ndiaye [2014] for $H > 1/2$. The strong consistency of the estimator is proved for $\theta < 0$ in Cénac and Es-Sebaiy [2015] and for $\theta > 0$ in Es-Sebaiy and Ndiaye [2014]. In Kubilius et al. [2015], they show the consistency for an estimator that is similar in form to the the maximum likelihood estimator obtained when the process is driven by a standard Brownian motion. Strong consistency is proved for $\theta > 0$ and consistency for $\theta \leq 0$ when $H \in (0,1)$. For

$H > 1/2$ the result is deduced directly from the previous works of Cénac and Es-Sebaiy [2015] and Es-Sebaiy and Ndiaye [2014].

For other estimators of the drift with discrete observation for $H > 1/2$ see Hu and Song [2013], Xiao et al. [2011]. Minimum contrast estimators for continuous and discrete times are considered in Bishwal [2011]. In Es-Sebaiy et al. [2019] least squares type estimators are considered for discrete observations of Gaussian Ornstein-Uhlenbeck processes that include the fractional case, for $\theta > 0$ with $H \in (0, 1)$.

Estimation of the drift parameter in the Volterra Ornstein-Uhlenbeck process (4.1) is a challenging problem. Indeed, this process is neither a semimartingale nor Markovian as in the fractional Ornstein-Uhlenbeck model. But, unlike in the fractional case, the kernel also appears next the parameter we want to estimate. This makes the statistical inference even more cumbersome.

This chapter is organized as follows. In Section 4.1 we compute the variance parameter of the Volterra Ornstein-Uhlenbeck process. Section 4.2 is dedicated to the study of the maximum likelihood estimator of the drift parameter when the process is observed in continuous time. Theorem 4.2.1 shows the strong consistency of this estimator when $\theta > 0$. In Section 4.3, we study the estimation of the drift parameter in the Volterra Ornstein-Uhlenbeck process based on discrete observations of the process. We construct an estimator that is similar in form to the maximum likelihood estimator and we show its weak consistency in Theorem 4.3.1. Finally, in Section 4.4, we test our estimators based on simulations of the process.

4.1 Estimation of the variance

Estimation of the variance parameter σ in (4.1) is a relatively easy task. Introduce the auxiliary semimartingale process Z defined by

$$Z_t := (L * X)(t) = \theta \int_0^t X_s ds + \sigma W_t, \quad t \geq 0, \tag{4.4}$$

where $L(t) = \frac{t^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}$ is the resolvent of the first kind of the kernel K (see Section 1.1 for a better understanding of stochastic convolutions and resolvents). The parameter σ^2 can be computed from the quadratic variation of the process Z in (4.4) on any finite time interval since

$$\begin{aligned} \sigma^2 t = \langle Z \rangle_t &= \lim_{n\to\infty} \sum_{k=1}^{n} (Z_{t_{k+1}} - Z_{t_k})^2, \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \big((L * X)(t_{k+1}) - (L * X)(t_k) \big)^2 \end{aligned}$$

where (t_k) is an appropriate partition of $[0, t]$ such that $\sup_k |t_{k+1} - t_k|$ tends to zero as n goes to infinity. Therefore, without loss of generality, in the following we assume that σ is known and equal to 1.

We are now interested in the estimation of the parameter θ in equation (4.4) and will define and study the maximum likelihood estimator of θ in the next section.

4.2 Maximum likelihood estimator of θ

The maximum likelihood estimation (refer to Liptser and Shiryaev [2001]) for θ is given by

$$\hat{\theta}_t = \frac{\int_0^t X_s dZ_s}{\int_0^t X_s^2 ds}, \quad t \geq 0. \tag{4.5}$$

It follows from the dynamic of Z that the estimation $\hat{\theta}_t$ of θ can be expressed as:

$$\hat{\theta}_t = \theta + \frac{\int_0^t X_s dW_s}{\int_0^t X_s^2 ds}. \tag{4.6}$$

Since $\hat{\theta}_t - \theta = \frac{M_t}{\langle M \rangle_t}$ where M is the martingale $M = \int_0^\cdot X_s dW_s$. The strong consistency of $\hat{\theta}_t$ is equivalent to the almost sure convergence of the process $\langle M \rangle = \int_0^\cdot X_s^2 ds$ to infinity, because of the strong law of large numbers.

Observe that the solution to equation (4.1) (see Abi Jaber et al. [2019, Lemma 2.5]) is the centered Gaussian process defined by

$$X_t = \int_0^t K_\theta(t-s) dW_s, \quad t \geq 0, \tag{4.7}$$

where K_θ is the function defined by

$$K_\theta(t) = t^{\alpha-1} E_{\alpha,\alpha}(\theta t^\alpha), \quad t \geq 0$$

and $E_{\alpha,\alpha}$ denotes the Mittag-Leffler function. Recall that given $\alpha, \beta > 0$, $E_{\alpha,\beta}$ is defined on $\mathbb{C}$ by

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}.$$

Theorem 4.2.1. *Let $\theta \geq 0$. The estimator $\hat{\theta}_t$ in (4.5) is strongly consistent.*

Proof. We focus on the Laplace transform of the process $\langle M \rangle$. It is sufficient to show that:

$$\forall w > 0, \quad \mathbb{E}\Big[e^{-w\int_0^t X_s^2 ds}\Big] \to 0 \quad \text{as } t \text{ goes to } +\infty. \tag{4.8}$$

Notice that for $w > 0$:

$$\mathbb{E}\left[\exp\left(-w\int_0^t X_s^2 ds\right)\right] \leq \mathbb{E}\left[\exp\left(-\frac{w}{t}\left(\int_0^t X_s ds\right)^2\right)\right],$$

and that for a random Gaussian variable ξ with mean m and variance σ^2, we have:

$$\mathbb{E}[\exp(-w\xi^2)] = (2w\sigma^2+1)^{-1/2}\exp\left(-\frac{wm^2}{2w\sigma^2+1}\right) \leq (2w\sigma^2+1)^{-1/2}.$$

Therefore, to prove (4.8) it is sufficient to exhibit that:

$$\mathrm{Var}\left(\frac{1}{\sqrt{t}}\int_0^t X_s\, ds\right) \to +\infty \text{ as } t \text{ goes to } +\infty.$$

As $\mathbb{E}\Big[\int_0^t X_s\, ds\Big] = 0$, we can conclude the proof, thanks to the next lemma, by asserting that:

$$\mathbb{E}\left[\frac{1}{t}\left(\int_0^t X_s\, ds\right)^2\right] \to +\infty \text{ as } t \text{ goes to } +\infty.$$

□

Lemma 4.2.2. *Let $T > 0$ we have the following lower bounds:*

i) If $\theta \geq 0$, we have

$$\mathbb{E}\left[\left(\int_0^T X_t\, dt\right)^2\right] \geq \frac{2T^{2\alpha+1}}{\Gamma(\alpha)\Gamma(\alpha+1)}\int_0^1 u(1-u)^{2\alpha-1}du.$$

ii) If $\theta < 0$, we have

$$\mathbb{E}\left[\left(\int_0^T X_t\, dt\right)^2\right] \geq \frac{T}{\theta^2} - \frac{2T^{1-\alpha}}{\theta^2\Gamma(1+\alpha)^{-1}|\theta|(1-\alpha)}.$$

Proof. Assume that $\theta > 0$, from (4.1) we get:

$$\int_0^T X_t\, dt = \int_0^T K(T-t)[W_t + \theta Y_t]dt,$$

with $Y_t = \int_0^t X_u\, du$. Then,

$$\left(\int_0^T X_t\,dt\right)^2 = \int_0^T\int_0^T K(T-u)W_uK(T-v)W_v dudv + \int_0^T\int_0^T K(T-u)K(T-v)\big[\theta^2 Y_uY_v + \theta W_uY_v + \theta W_vY_u\big]dudv. \quad (4.9)$$

Hence,

$$\begin{aligned}\mathbb{E}[Y_uY_v] &= \mathbb{E}\Big[\Big(\int_0^u X_s ds\Big)\Big(\int_0^v X_r dr\Big)\Big]\\ &= \mathbb{E}\Big[\Big(\int_0^u K_\theta(u-s)W_s ds\Big)\Big(\int_0^v K_\theta(v-r)W_r dr\Big)\Big]\\ &= \int_0^u\int_0^v K_\theta(u-s)K_\theta(v-r)\mathbb{E}[W_uW_v]drds,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[Y_vW_u] &= \mathbb{E}\Big[\Big(\int_0^v K_\theta(v-r)W_r dr\Big)W_u\Big]\\ &= \int_0^v K_\theta(v-r)\mathbb{E}[W_rW_u]dr.\end{aligned}$$

Since for any $u, v > 0$ $\mathbb{E}[W_uW_v] = \min(u, v)$ and $K_\theta(u)$ is positive for every $u > 0$, we have $\mathbb{E}[Y_uY_v] > 0$ and $\mathbb{E}[Y_vW_u] > 0$. Hence, taking expectation in (4.9) we obtain,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T X_t\,dt\right)^2\right] &\geq \int_0^T\int_0^T K(T-u)K(T-v)\mathbb{E}[W_uW_v]dudv\\ &= \frac{2}{\Gamma(\alpha)^2}\int_0^T (T-v)^{\alpha-1}v\int_v^T (T-u)^{\alpha-1}dudv\\ &= \frac{2}{\Gamma(\alpha)^2\alpha}\int_0^T v(T-v)^{2\alpha-1}dv\\ &= \frac{2T^{2\alpha+1}}{\Gamma(\alpha)\Gamma(\alpha+1)}\int_0^1 u(1-u)^{2\alpha-1}du\\ &= \frac{2T^{2\alpha+1}}{\Gamma(\alpha)\Gamma(\alpha+1)}\beta(2,2\alpha),\end{aligned}$$

where β is the beta function.

Then, the variance can be bounded from below by the value:

$$\mathbb{E}\left[\frac{1}{T}\left(\int_0^T X_t\,dt\right)^2\right] \geq \frac{2T^{2\alpha}}{\Gamma(\alpha)\Gamma(\alpha+1)}\beta(2,2\alpha).$$

Observe that $\alpha > 0$, therefore the last expression goes to infinity as T goes to infinity.

Now, assume that $\theta < 0$, from (4.7):

$$\int_0^T X_t\, dt = \left((1 * \mathcal{E}_\theta) * dW\right)(T).$$

Hence,

$$\mathbb{E}\left[\left(\int_0^T X_t\, dt\right)^2\right] = \int_0^T \left((1 * \mathcal{E}_\theta)(t)\right)^2\, dt.$$

Yielding to the following inferior bound for the variance:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T X_t\, dt\right)^2\right] &= \int_0^T \frac{1}{\theta}(E_{\alpha,1}(\theta t^\alpha) - 1)^2 dt \\
&= \frac{1}{\theta^2}\int_0^T \left(E_{\alpha,1}^2(\theta t^\alpha) - 2E_{\alpha,1}(\theta t^\alpha) + 1\right)\, dt \\
&\geq \frac{T}{\theta^2} - \frac{2}{\theta^2}\int_0^T E_{\alpha,1}(\theta t^\alpha) dt. \\
&\geq \frac{T}{\theta^2} - \frac{2}{\theta^2}\int_0^T \frac{1}{1 + \Gamma(1+\alpha)^{-1}(|\theta| t^\alpha)} dt. \\
&\geq \frac{T}{\theta^2} - \frac{2}{\theta^2}\int_0^T \frac{1}{\Gamma(1+\alpha)^{-1}(|\theta| t^\alpha)} dt. \\
&= \frac{T}{\theta^2} - \frac{2T^{1-\alpha}}{\theta^2 \Gamma(1+\alpha)^{-1}|\theta|(1-\alpha)},
\end{aligned}$$

□

Remark 4.2.3. Observe that $0 < 1 - \alpha < 1/2$. Therefore, even when $\theta < 0$, the variance of $\int_0^T X_s\, ds$ goes to infinity as T goes to infinity, however it does not imply strong consistency for the continuous estimator $\hat{\theta}$. Indeed, if we follow the proof of Theorem 4.2.1, Jensen inequality implies that we need that the variance of $\frac{1}{\sqrt{T}}\int_0^T X_s\, ds$ goes to infinity as T goes to infinity. Unfortunately, it is not the case when $\theta < 0$ and the asymptotic behavior of the Laplace transform of $\int_0^T X_s^2\, ds$ should be directly studied. A way to do so could be to exploit the following result of Abi Jaber [2019b] :

$$\mathbb{E}\left[\exp\left(-\int_0^T wX_t^2\, dt\right)\right] = \lim_{n\to\infty} \frac{1}{\det\left(\mathbb{I}_n + \frac{2Tw}{n} C_0^n\right)^{1/2}}$$

where C_0^n represents the covariance matrix $C_0(s_i^n, s_k^n)$ for $i, k = 1, \dots, n$ with $s_i^n = \frac{iT}{n}$ for $i = 1, \dots, n$.

For $0 \leq s, u \leq T$

$$C_0(s,u) = \int_0^{u \wedge s} (s-z)^{\alpha-1} E_{\alpha,\alpha}(\theta(s-z)^\alpha)(u-z)^{\alpha-1} E_{\alpha,\alpha}(\theta(u-z)^\alpha) dz.$$

4.3 Estimator of θ for discrete observations of the process X

Following Kubilius et al. [2015], we assume that the process X is observed at discrete times with interval time between observations given by n^{-1}, where $n \geq 1$. We set $t_k = \frac{k}{n}$, $0 \leq k \leq a_n := nb_n$, with a and $b \in \mathbb{N}^{\mathbb{N}}$ such that $\lim_{n \to +\infty} b_n = +\infty$. The process X is observed at points $\{t_k : 0 \leq k \leq a_n\}$.
Notice that when n goes to infinity, the interval time between observations tends to zero, which means that we work with high-frequency data. Also, the number of observation increases to infinity with speed a_n. Inspired by (4.5), we consider the following estimator

$$\bar{\theta}_n(a) = \frac{\sum_{k=0}^{a_n-1} X_{t_k} \Delta Z_{t_k}}{\frac{1}{n} \sum_{k=0}^{a_n-1} X_{t_k}^2}, \tag{4.10}$$

where $\Delta Z_{t_k} = Z_{t_{k+1}} - Z_{t_k}$. From (4.4), the estimator $\bar{\theta}_n(a)$ satisfies the following equality

$$\bar{\theta}_n(a) \quad = \quad \theta + \frac{\theta \sum_{k=0}^{a_n-1} X_{t_k} \int_{\frac{k}{n}}^{\frac{k+1}{n}} (X_s - X_{t_k}) ds + \sum_{k=0}^{a_n-1} X_{t_k} \Delta W_{t_k}}{\frac{1}{n} \sum_{k=0}^{a_n-1} X_{t_k}^2}. \tag{4.11}$$

Following Kubilius et al. [2015], in which the consistency for a discrete estimator of the drift coefficient of a fractional Ornstein-Uhlenbeck process is proved, we establish in this section the weak consistency of the estimator (4.11) .

Theorem 4.3.1. ***Weak consistency of the discrete estimator***
Define $\delta := \frac{2\alpha-1}{2}$ and $\gamma := \mid \theta \mid^{\frac{1}{\alpha}}$. For any $\theta \in \mathbb{R}$, if there exists $0 < \beta$ such that $\gamma\beta < \frac{\delta}{2}$ and

$$b_n := \frac{a_n}{n} \leq \beta \ln(n), \quad n \geq 1.$$

then the sequence $(\bar{\theta}_n(a))_{n \geq 0}$ is weakly consistent i.e. $(\bar{\theta}_n(a))_{n \geq 0}$ converges to θ in probability as $n \to \infty$.
Moreover, if $\theta > 0$ and there exists $0 < \beta < \frac{\delta}{2\delta + \frac{7}{2}}$ such that

$$b_n \leq n^\beta, \quad n \geq 1;$$

then the sequence $(\bar{\theta}_n(a))_{n \geq 0}$ is also weakly consistent

We start with some expansions of the variance of the Gaussian process $(X_t)_{t\geq 0}$. From integral representations of Mittag-Leffler functions, expansions for $\mid z \mid$ going to $+\infty$ are obtained in Gorenflo et al. [2002]. In particular, for $x \in \mathbb{R}$, we have

$$E_{\alpha,\beta}(x) = \begin{cases} \frac{1}{\alpha} x^{\frac{1-\beta}{\alpha}} e^{x^{\frac{1}{\alpha}}} + \mathrm{o}(1), & \text{for } x \to +\infty \\ \\ -\frac{1}{\Gamma(\beta-\alpha)x} + \mathrm{O}(\frac{1}{x^2}), & \text{for } x \to -\infty \text{ and } \alpha \neq \beta. \end{cases} \tag{4.12}$$

Therefore, we have the following estimations for the variance of $(X_t)_{t\geq 0}$.

Lemma 4.3.2. ***Variance estimation***

For t going to $+\infty$ and $\theta \neq 0$, we have

$$\mathbb{E}\left[X_t^2\right] \leq \frac{t}{\alpha^2(2\alpha-1)} \mid \theta \mid^{\frac{2(1-\alpha)}{\alpha}} e^{2|\theta|^{\frac{1}{\alpha}} t} + o(t^{\alpha} e^{2|\theta|^{\frac{1}{\alpha}} t}), \tag{4.13}$$

For $\theta = 0$ we have $\mathbb{E}[X_t^2] = \frac{t^{2\alpha-1}}{(2\alpha-1)\Gamma(\alpha)^2}$.

Proof. Recall that X is defined, for $t \geq 0$, by

$$X_t = \int_0^t K_\theta(t-s)dW_s, \quad \text{with } K_\theta(u) := u^{\alpha-1}E_{\alpha,\alpha}(\theta u^{\alpha}).$$

Hence, we have

$$\mathbb{E}\Big[X_t^2\Big] = \int_0^t K_\theta^2(t-s)\, ds, \quad t \geq 0, \tag{4.14}$$

and

$$\begin{aligned} \mathbb{E}\Big[X_t^2\Big] &= \int_0^t K_\theta^2(t-s)\, ds \\ &= \sum_{k,n=0}^{+\infty} \frac{\theta^{n+k}}{\Gamma(\alpha(n+1))\Gamma(\alpha(k+1))} \int_0^t u^{(n+k+2)\alpha-2}\, ds \\ &= \sum_{k,n=0}^{+\infty} \frac{\theta^{n+k}}{\Gamma(\alpha(n+1))\Gamma(\alpha(k+1))} \frac{t^{(n+k+2)\alpha-1}}{(n+k+2)\alpha-1}. \end{aligned}$$

For $n, k \in \mathbb{N}$, as $\frac{1}{2} < \alpha \leq 1$, we have

$$0 < 2\alpha - 1 \leq (n+k+2)\alpha - 1 \leq 2\alpha(n \vee k + 1).$$

Taking the absolute value, we get

$$\begin{aligned} \mathbb{E}\Big[X_t^2\Big] &\leq \frac{t^{2\alpha-1}}{2\alpha-1} \sum_{k,n=0}^{+\infty} \frac{(\mid \theta \mid t^{\alpha})^{n+k}}{\Gamma(\alpha(n+1))\Gamma(\alpha(k+1))} \\ &= \frac{t^{2\alpha-1}}{2\alpha-1} \left(E_{\alpha,\alpha}(\mid \theta \mid t^{\alpha})\right)^2. \end{aligned}$$

From the estimation (4.12), it follows that if $\theta \neq 0$, we get

$$\mathbb{E}\Big[X_t^2\Big] = \frac{t}{\alpha^2(2\alpha-1)} \mid \theta \mid^{\frac{2(1-\alpha)}{\alpha}} e^{2|\theta|^{\frac{1}{\alpha}} t} + \mathrm{o}(t^{\alpha} e^{2|\theta|^{\frac{1}{\alpha}} t}), \quad \text{for } t \to +\infty,$$

and, if $\theta = 0$, as $E_{\alpha,\alpha}(0) = \Gamma(\alpha)^{-1}$, we have

$$\mathbb{E}\Big[X_t^2\Big] = \frac{t^{2\alpha-1}}{(2\alpha-1)\Gamma(\alpha)^2}.$$

□

To prove Theorem 4.3.1, we need to control X and its increments. We do this in the next lemmas.

Lemma 4.3.3. *There exists a non-negative random variable ξ and a constant $c(\xi) > 0$ such that: $\mathbb{E}[e^{y\xi^2}] < +\infty$ for any $0 < y < c(\xi)$, and for $t > 0$*

$$S_t := \sup_{0 \le s \le t} \mid X_s \mid \le \Big(\Big(t^{\delta} \ln^2(t)\Big) \vee 1\Big)\, \xi (e^{\gamma t} + 1), \quad \textit{with } \gamma := \mid \theta \mid^{\frac{1}{\alpha}} \ \textit{ and } \delta := \frac{2\alpha - 1}{2}.$$

Proof. For $t > 0$, we have

$$\begin{aligned}
S_t &\le \sup_{0 \le s \le t} \mid \theta \mid \int_0^s K(s-u) \mid X_u \mid du + \left| \int_0^s K(s-u) dW_u \right| \\
&\le \mid \theta \mid \sup_{0 \le s \le t} \int_0^s K(v) S_{s-v}\, dv + \sup_{0 \le s \le t} \left| \int_0^s K(s-u) dW_u \right| \\
&\le \mid \theta \mid \int_0^t K(v) S_{t-v}\, dv + \sup_{0 \le s \le t} \left| \int_0^s K(s-u) dW_u \right| \\
&\le \mid \theta \mid \int_0^t K(t-u) S_u\, du + \sup_{0 \le s \le t} \left| \int_0^s K(s-u) dW_u \right|
\end{aligned}$$

From Kubilius et al. [2015, Proposition 3.1], for any $p > 1$, there exists a non-negative random variable $\hat{\xi}$ and a constant $c(\hat{\xi}) > 0$ such that: $\mathbb{E}[e^{y\hat{\xi}^2}] < +\infty$ for any $0 < y < c(\hat{\xi})$, and

$$\sup_{0 \le s \le t} \left| \int_0^s K(s-u) dW_u \right| \le \Big(\Big(t^{\delta} \mid \ln(t) \mid^2\Big) \vee 1\Big)\, \hat{\xi}, \quad t > 0.$$

Hence, we get

$$S_t \le \mid \theta \mid \int_0^t K(t-u) S_u\, du + \Big(\Big(t^{\delta} \mid \ln(t) \mid^2\Big) \vee 1\Big)\, \hat{\xi} \tag{4.15}$$

From an extension of Grönwall's inequality to Volterra processes (see in Ye et al. [2007, Corollary 2]), it follows that

$$S_t \le \Big(\Big(t^{\delta} \ln^2(t)\Big) \vee 1\Big)\, \hat{\xi} E_{\alpha,1}(\mid \theta \mid t^{\alpha}).$$

Moreover, the asymptotic expansion of the Mittag-Leffler function $E_{\alpha,1}$ when z goes to $+\infty$ is given by

$$E_{\alpha,1}(z) = \frac{e^{z^{\frac{1}{\alpha}}}}{\alpha} + \varepsilon(z), \quad \text{with } \varepsilon \text{ continuous on } \mathbb{R}^+ \text{ and } \lim_{z\to+\infty} \varepsilon(z) = 0.$$

We refer to Gorenflo et al. [2002] but it is a well-known consequence of integral representation of the Mittag-Leffler function. We can then find a constant $c > 0$ such that $\xi = c\hat{\xi}$ and

$$S_t \leq \left(\left(t^\delta \ln^2(t)\right) \vee 1\right) \xi(e^{|\theta|^{\frac{1}{\alpha}} t} + 1).$$

□

Lemma 4.3.4. *For any $0 < \varepsilon < \delta$ and a non-negative random variable η and a constant $c(\eta) > 0$ such that: $\mathbb{E}[e^{y\eta^2}] < +\infty$ for any $0 < y < c(\eta)$, and for any $k \in [0, a_n - 1]$, and $t \in [t_k, t_{k+1})$, we have*

$$\sup_{t_k\leq s\leq t} | X_s - X_{t_k} | \leq C(t-t_k)^{\delta-\varepsilon}\left[\left(| \ln(t-t_k) |^{\frac{1}{2}} +1\right)\ln(t+2)\eta + S_t(t-t_k)^{\alpha-\delta+\varepsilon} + \left(\int_0^t X_u^2\, du\right)^{\frac{1}{2}} (t-t_k)^\varepsilon\right].$$

Proof. Let $t > t_k$, we set $\Delta S_t^k = \sup_{t_k\leq s\leq t} | X_s - X_{t_k} |$ and $M_t = \int_0^t K(t-u)\, dW_u$. Mainly from Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\Delta S_t^k &\leq \sup_{t_k\leq s\leq t} | \theta | \left| \int_0^s K(s-u)X_u du - \int_0^{t_k} K(t_k-u)X_u du \right| + | M_s - M_{t_k} | \\
&\leq | \theta | \sup_{t_k\leq s\leq t} \int_{t_k}^s K(s-u) | X_u | du + \int_0^{t_k} | K(s-u) - K(t_k-u) || X_u | du + \sup_{t_k\leq s\leq t} | M_s - M_{t_k} | \\
&\leq | \theta | S_t\sqrt{t-t_k} \left(\int_0^{t-t_k} K^2(v)\, dv\right)^{\frac{1}{2}} \\
&\quad + \left(\int_0^{t_k} X_u^2\, du\right)^{\frac{1}{2}} \sup_{t_k\leq s\leq t} \left(\int_0^{t_k} (K(v+s-t_k) - K(v))^2\, dv\right)^{\frac{1}{2}} \\
&\quad + \sup_{t_k\leq s\leq t} | M_s - M_{t_k} |
\end{aligned}$$

From the definition of the fractional kernel, there exists $C > 0$ such that

$$\int_0^{t-t_k} K^2(v)\, dv = C(t-t_k)^{2\alpha-1} \quad \text{and} \quad \int_0^{t_k} (K(v+s-t_k)-K(v))^2\, dv \leq C(t-t_k)^{2\alpha-1}, \ \forall t_k \leq s \leq t.$$

From Kubilius et al. [2015, Proposition 3.1], there exists a non-negative random variable η such that there exists $c(\eta) > 0$: $\mathbb{E}[e^{y\eta^2}] < +\infty$ for any $0 < y < c(\eta)$, and such that for all $t > t_k$,

$$\begin{aligned}
\sup_{t_k\leq s\leq t} | M_s - M_{t_k} | &\leq \sup_{t_k\leq s\leq t} (s-t_k)^\delta \left(| \ln(s-t_k) |^{\frac{1}{2}} +1\right) \ln(s+2)\eta, \\
&\leq C(t-t_k)^{\delta-\varepsilon} \left(| \ln(t-t_k) |^{\frac{1}{2}} +1\right) \ln(t+2)\eta,
\end{aligned}$$

for any $\varepsilon \in (0, \delta)$ and with a constant $C > 0$ depending on ε. Hence, for any $\varepsilon \in (0, \delta)$, there exists a generic constant $C > 0$ such that

$$
\begin{aligned}
\Delta S_t^k &\leq CS_t(t-t_k)^{\alpha} + C\left(\int_0^t X_u^2\,du\right)^{\frac{1}{2}}(t-t_k)^{\delta} + C(t-t_k)^{\delta-\varepsilon}\left(|\ln(t-t_k)|^{\frac{1}{2}}+1\right)\ln(t+2)\eta \\
&= C(t-t_k)^{\delta-\varepsilon}\left[S_t(t-t_k)^{\alpha-\delta+\varepsilon} + \left(\int_0^t X_u^2\,du\right)^{\frac{1}{2}}(t-t_k)^{\varepsilon} + [\left(|\ln(t-t_k)|^{\frac{1}{2}}+1\right)\ln(t+2)\eta\right]
\end{aligned}
$$

Notice that if $\theta = 0$, we have

$$
\Delta S_t^k \leq \sup_{t_k \leq s \leq t} | M_s - M_{t_k} | \leq C(t-t_k)^{\delta-\varepsilon}\left(|\ln(t-t_k)|^{\frac{1}{2}}+1\right)\ln(t+2)\eta.
$$

Hence our result is true for any $\theta \in \mathbb{R}$. □

We will now apply the two previous lemmas to get an upper bound for the numerator of the fraction which appears in equation (4.11). First, it follows from Lemma 4.3.3 that

$$
\begin{aligned}
\int_0^t X_u^2\,du &\leq \int_0^t S_u^2\,du \\
&\leq \xi^2\int_0^t \left(\left(u^{2\delta}\ln^4(u)\right)\vee 1\right)(e^{\gamma u}+1)^2\,du \\
&\leq C\xi^2(e^{\gamma t}+1)^2(t^{2\delta+1+\varepsilon}+1).
\end{aligned}
$$

Therefore, we deduce, from Lemmas 4.3.3 and 4.3.4, that, for $t_k \leq t \leq t_{k+1}$, and any $\varepsilon \in (0, \frac{\delta}{2})$

$$
\begin{aligned}
\Delta S_t^k &\leq C\frac{1}{n^{\delta+1/2}}S_t + \frac{C}{n^{\delta}}\left(\int_0^t X_u^2\,du\right)^{\frac{1}{2}} + C\frac{1}{n^{\delta-\varepsilon}}\left(\ln(n)^{\frac{1}{2}}+1\right)\ln(t+2)\eta \\
&\leq \frac{C}{n^{\delta}}\xi(e^{\gamma t}+1)\left(\frac{1}{n^{1/2}}t^{\delta+\varepsilon} + t^{\delta+\frac{1+\varepsilon}{2}}+1\right) + C\frac{1}{n^{\delta-\varepsilon}}\left(\ln(n)^{\frac{1}{2}}+1\right)\ln(t+2)\eta \\
&\leq \frac{C}{n^{\delta-2\varepsilon}}\xi(e^{\gamma t}+1)\left(t^{\delta+\frac{1+\varepsilon}{2}}+1\right),
\end{aligned}
\tag{4.16}
$$

where $C > 0$ is a generic constant which , in particular, does not depend on k, n and t.

Lemma 4.3.5. *For any $n \geq 1$, recall that we set $b_n := \frac{a_n}{n}$. For any $\varepsilon \in (0, \frac{\delta}{2})$, there exists a constant $C > 0$ such that*

$$
\left|\sum_{k=0}^{a_n-1} X_{t_k}\int_{\frac{k}{n}}^{\frac{k+1}{n}}(X_s - X_{t_k})ds\right| \leq \frac{C\xi^2}{n^{\delta-2\varepsilon}}b_n^{2\delta+\frac{3}{2}(\varepsilon+1)}e^{2\gamma b_n}.
$$

Proof. For $n > 0$, we apply Lemma 4.3.3 and equation (4.16) to get

$$\begin{aligned}
\left|\sum_{k=0}^{a_n-1} X_{t_k}\int_{\frac{k}{n}}^{\frac{k+1}{n}}(X_s - X_{t_k})ds\right| &\leq \frac{1}{n}\sum_{k=0}^{a_n-1} S_{t_k}\Delta S^k_{t_k+1}\\
&\leq \frac{C\xi}{n}\sum_{k=0}^{a_n-1}\left[\left(\left(t_k^\delta \ln^2(t_k)\right)\vee 1\right)(e^{\gamma t_k}+1)\right]\left[\frac{1}{n^{\delta-2\varepsilon}}\xi(e^{\gamma t_{k+1}}+1)\left({t_{k+1}}^{\delta+\frac{1+\varepsilon}{2}}+1\right)\right]\\
&\leq \frac{C\xi^2 e^{2\gamma b_n}}{n^{\delta-2\varepsilon+1}}\sum_{k=0}^{a_n-1} t_{k+1}^{2\delta+\frac{3\varepsilon}{2}+\frac{1}{2}}\\
&\leq \frac{C\xi^2}{n^{\delta-2\varepsilon}} b_n^{2\delta+\frac{3\varepsilon}{2}+\frac{3}{2}} e^{2\gamma b_n},
\end{aligned}$$

where $C > 0$ is a generic positive real number which does not depend on k and n. □

In the next lemma, we deal with the second term of the numerator in equation (4.11).

Lemma 4.3.6. *For any $n \geq 1$, there exists a random variables ξ_n^+ such that*

$$\sup_{n\geq 1}\mathbb{E}[(\xi_n^+)^2] < +\infty, \qquad \textit{and} \qquad \sum_{k=0}^{a_n-1} X_{t_k}\Delta W_{t_k} = \sqrt{b_n}e^{\gamma b_n}\xi_n^+.$$

Proof. Let $n \geq 0$. From equation (4.7), we have

$$\mathbb{E}\left[\left(\sum_{k=0}^{a_n-1} X_{t_k}\Delta W_{t_k}\right)^2\right] \leq \sum_{k,j=0}^{a_n-1}\mathbb{E}\left[\int_0^{t_k} K_\theta(t_k-u)\,dW_u \int_0^{t_j} K_\theta(t_j-u)\,dW_u\Delta W_{t_k}\Delta W_{t_j}\right].$$

As the vector $\left(\int_0^{t_k} K_\theta(t_k-u)\,dW_u,\ \int_0^{t_j} K_\theta(t_j-u)\,dW_u,\ \Delta W_{t_k},\ \Delta W_{t_j}\right)$ is gaussian for any $j,k \in \mathbb{N}$, it follows from Isserlis' formula and the independence between Brownian increments that

$$\mathbb{E}\Big[\int_0^{t_k} K_\theta(t_k-u)\,dW_u \int_0^{t_j} K_\theta(t_j-u)\,dW_u\Delta W_{t_k}\Delta W_{t_j}\Big] = \mathbb{1}_{\{j=k\}}\mathbb{E}\Big[\Delta W_{t_k}^2\Big]\mathbb{E}\Big[\Big(\int_0^{t_k} K_\theta(t_k-u)\,dW_u\Big)^2\Big].$$

Hence, we have

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{a_n-1} X_{t_k}\Delta W_{t_k}\Big)^2\Big] \leq \sum_{k=0}^{a_n-1}(t_{k+1}-t_k)\mathbb{E}\Big[\Big(\int_0^{t_k} K_\theta(t_k-u)\,dW_u\Big)^2\Big] = \frac{1}{n}\sum_{k=0}^{a_n-1}\mathbb{E}\Big[X_{t_k}^2\Big]$$

. Lemma 4.13 implies that there exists a constant $C > 0$ such that

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{a_n-1} X_{t_k}\Delta W_{t_k}\Big)^2\Big] \leq \frac{C}{\alpha^2(2\alpha-1)n}\sum_{k=0}^{a_n-1} t_k e^{2\gamma t_k} \leq \frac{C}{\alpha^2(2\alpha-1)} b_n e^{2\gamma b_n} \tag{4.17}$$

For $n \geq 1$, we set

$$\xi_n^+ := b_n^{-\frac{1}{2}} e^{-\gamma b_n}\sum_{k=0}^{a_n-1} X_{t_k}\Delta W_{t_k}$$

We can conclude the proof by asserting that equations (4.17) gives

$$\sup_{n\geq 1}\mathbb{E}[(\xi_n^+)^2] \leq \frac{C}{\alpha^2(2\alpha-1)} < +\infty. \tag{4.18}$$

□

4.3.1 Proof of Theorem 4.3.1

We conclude this section with the proof of the consistency of the estimator $\bar{\theta}_n(a)$.

From Lemmas 4.3.5 and 4.3.6, for $n \geq 1$, we have

$$
\begin{aligned}
\mid \bar{\theta}_n(a) - \theta \mid &= \left| \frac{\theta \sum_{k=0}^{a_n-1} X_{t_k} \int_{\frac{k}{n}}^{\frac{k+1}{n}} (X_s - X_{t_k}) ds + \sum_{k=0}^{a_n-1} X_{t_k} \Delta W_{t_k}}{\frac{1}{n} \sum_{k=0}^{a_n-1} X_{t_k}^2} \right| \\
&\leq \frac{C \mid \theta \mid \xi^2 e^{2\gamma b_n} b_n^{2\delta + \frac{3\varepsilon}{2} + \frac{3}{2}} n^{2\varepsilon - \delta} + \sqrt{b_n} e^{\gamma b_n} \xi_n^+}{\frac{1}{n} \sum_{k=0}^{nb_n-1} X_{t_k}^2}. \qquad (4.19)
\end{aligned}
$$

We now have to find a lower bound for the denominator. Notice that we have

$$
\begin{aligned}
\left| \int_0^{b_n} X_s^2 \, ds - \frac{1}{n} \sum_{k=0}^{nb_n-1} X_{t_k}^2 \right| &\leq \sum_{k=0}^{nb_n-1} \int_{t_k}^{t_{k+1}} \mid X_s^2 - X_{t_k}^2 \mid ds \\
&\leq \sum_{k=0}^{nb_n-1} \int_{t_k}^{t_{k+1}} \mid X_s - X_{t_k} \mid (\mid X_s \mid + \mid X_{t_k} \mid) \, ds \\
&\leq \frac{2}{n} \sum_{k=0}^{nb_n-1} \Delta S_{t_{k+1}}^k S_{t_{k+1}}.
\end{aligned}
$$

We follow the proof of Lemma 4.3.5 and apply Lemma 4.3.3 and equation (4.16) to deduce that for any $0 < \varepsilon < \frac{\delta}{2}$, there exists $C > 0$ such that

$$
\left| \int_0^{b_n} X_s^2 \, ds - \frac{1}{n} \sum_{k=0}^{nb_n-1} X_{t_k}^2 \right| \leq \frac{C\xi^2 e^{2\gamma b_n}}{n^{\delta - 2\varepsilon}} b_n^{2\delta + \frac{3\varepsilon}{2} + \frac{3}{2}}.
$$

Hence, for $0 < \varepsilon < \frac{\delta}{2}$

$$
\frac{1}{n} \sum_{k=0}^{nb_n-1} X_{t_k}^2 = \int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon)
$$

where

$$
|\vartheta_n(\varepsilon)| \leq \frac{C\xi^2 e^{2\gamma b_n}}{n^{\delta - 2\varepsilon}} b_n^{2\delta + \frac{3\varepsilon}{2} + \frac{3}{2}}. \qquad (4.20)
$$

Now, we prove that, for any $\varepsilon > 0$ small enough, the following ratios tends to 0 in probability:

$$
K_n^1(\varepsilon) := \frac{e^{2\gamma b_n} b_n^{2\delta + \frac{3\varepsilon}{2} + \frac{3}{2}} n^{2\varepsilon - \delta}}{\int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon)} \quad \text{and} \quad K_n^2(\varepsilon) := \frac{\sqrt{b_n} e^{\gamma b_n} \xi_n^+}{\int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon)}.
$$

Let $\varepsilon \in (0, \frac{\delta}{2})$. In order to proof that $K_n^1(\varepsilon) \to 0$ in probability as $n \to +\infty$ is is sufficient to prove that

$$
e^{-2\gamma b_n} b_n^{-2\delta - \frac{3\varepsilon}{2} - \frac{3}{2}} n^{\delta - 2\varepsilon} \int_0^{b_n} X_s^2 ds + e^{-2\gamma b_n} b_n^{-2\delta - \frac{3\varepsilon}{2} - \frac{3}{2}} n^{\delta - 2\varepsilon} \vartheta_n(\varepsilon) \to +\infty \text{ in probability as } n \to +\infty.
$$

In view of (4.20) it is equivalent to

$$e^{-2\gamma b_n} b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{3}{2}} n^{\delta-2\varepsilon} \int_0^{b_n} X_s^2 ds \to +\infty \text{ in probability as } n \to +\infty$$

It follows from Jensen inequality that

$$e^{-2\gamma b_n} b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{3}{2}} n^{\delta-2\varepsilon} \int_0^{b_n} X_s^2 ds \geq e^{-2\gamma b_n} b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{5}{2}} n^{\delta-2\varepsilon} \left(\int_0^{b_n} X_s ds \right)^2.$$

Note that $\int_0^{b_n} X_s ds$ is a Gaussian process with the mean 0 and variance $\sigma_n^2 = \mathbb{E}\Big[\Big(\int_0^{b_n} X_s ds \Big)^2 \Big]$. Let $c_n := e^{-2\gamma b_n} b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{5}{2}} n^{\delta-2\varepsilon}$ and denote by $\mathcal{N}(0,1)$ the standard Gaussian random variable and $\Phi(x)$ its density function. We can deduce that for any $A > 0$ we have

$$\begin{aligned}
\mathbf{P}\left\{ c_n \left(\int_0^{b_n} X_s ds \right)^2 \leq A^2 \right\} &= \mathbf{P}\left\{ c_n^{1/2} \left| \int_0^{b_n} X_s ds \right| \leq A \right\} \\
&= \mathbf{P}\left\{ c_n^{1/2} \left| \sigma_n \mathcal{N}(0,1) \right| \leq A \right\} \\
&= \Phi\left(\frac{A}{\sigma_n c_n^{1/2}} \right) - \Phi\left(\frac{-A}{\sigma_n c_n^{1/2}} \right) \\
&\leq 2\left(\frac{A}{\sigma_n c_n^{1/2}} \right).
\end{aligned}$$

Now, let find a lower bound for σ_n. From Lemma 4.2.2, we know that there exists a positive constant $\underline{\sigma} > 0$ such that $\sigma_n^2 \geq \underline{\sigma}^2 b_n$ for any n large enough. Let $0 < \beta$ such that $\gamma\beta < \frac{\delta}{2}$ and $0 < \varepsilon < \frac{1}{2}(\delta - 2\gamma\beta)$, we assume that

$$b_n \leq \beta \ln(n), \quad \text{for any } n \geq 1.$$

For n great enough, we get

$$\sigma_n^2 c_n \geq \underline{\sigma}^2 \left(\beta \ln(n)\right)^{-2\delta-\frac{3\varepsilon}{2}-\frac{3}{2}} n^{\delta-2\gamma\beta-2\varepsilon},$$

and therefore, as c_n tends to $+\infty$ when n goes to $+\infty$,

$$\lim_{n\to+\infty} \mathbf{P}\left\{ c_n \left(\int_0^{b_n} X_s ds \right)^2 \leq A^2 \right\} = 0.$$

Now, we prove the convergence of $K_n^2(\varepsilon)$ to 0 in probability as n goes to $+\infty$.

For any $A > 0$ and $x_n > 0$ we have,

$$\begin{aligned}
\mathbf{P}\left\{ K_n^2(\varepsilon) > A \right\} &= \mathbf{P}\left\{ \frac{\sqrt{b_n} e^{\gamma b_n} \xi_n^+}{\int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon)} > A \right\} \\
&\leq \mathbf{P}\left\{ \xi_n^+ > x_n \right\} + \mathbf{P}\left\{ b_n^{-\frac{1}{2}} e^{-\gamma b_n} \left(\int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon) \right) < \frac{x_n}{A} \right\}
\end{aligned}$$

Let $x_n := e^{\gamma b_n} b_n^{2\delta+\frac{3\varepsilon}{2}+1} n^{2\varepsilon-\delta}$. Applying Markov's inequality

$$\mathbf{P}\left\{K_n^2(\varepsilon) > A\right\} \quad \leq \quad \frac{\mathbb{E}\left[\xi_n^{+^2}\right]}{x_n^2} + \mathbf{P}\left\{e^{-2\gamma b_n} b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{3}{2}} n^{\delta-2\varepsilon}\left(\int_0^{b_n} X_s^2 ds + \vartheta_n(\varepsilon)\right) < \frac{1}{A}\right\}$$

Thanks to (4.18) and the study of $K_n^1(\varepsilon)$ we conclude the first part of the proof.

To improve our result, we now assume that $\theta > 0$ and find another lower bound for σ_n.

Lemma 4.3.7. *Assume that $\theta > 0$. For any $n \geq 1$, we have*

$$\sigma_n^2 \geq C b_n^{-1} \frac{1}{2\alpha^2 \theta^{2\frac{\alpha+1}{\alpha}}} e^{2\theta^{1/\alpha} b_n}$$

for some constant $C > 0$.

The proof of this lemma is detailed after the end of the proof of Theorem 4.3.1. Recall that $K_n^1(\varepsilon)$ tends to 0 in probability as soon as $\lim_{n\to+\infty} \sigma_n^2 c_n = +\infty$. Therefore, if we assume that $b_n \leq n^\beta$ for all $n \geq 1$ with $0 < \beta < \frac{\delta}{2\delta+\frac{7}{2}}$ then there exists $C > 0$ such that

$$\begin{aligned} \sigma_n^2 c_n \quad &\geq \quad C b_n^{-2\delta-\frac{3\varepsilon}{2}-\frac{7}{2}} n^{\delta-2\varepsilon} \\ &\geq \quad C n^{\delta-2\varepsilon-\beta(2\delta+\frac{3\varepsilon}{2}+\frac{7}{2})}. \end{aligned}$$

Hence, for any $0 < \varepsilon < \frac{1}{2+\frac{3\beta}{2}}\left(\delta - \beta(2\delta+\frac{7}{2})\right)$, $\lim_{n\to+\infty} \sigma_n^2 c_n = +\infty$. Following the previous steps, we can also prove that $K_n^2(\varepsilon)$ tends to 0 in probability when n goes to $+\infty$ and conclude the proof of Theorem 4.3.1. □

Proof of Lemma 4.3.7: Using equation (4.7) we have

$$Y_n := \int_0^{b_n} X_t dt = \left((1 * K_\theta) * dW\right)_{b_n},$$

with mean $\mathbb{E}[Y_n] = 0$ and variance σ_n^2 given by

$$\sigma_n^2 = \mathbb{E}\left[Y_n^2\right] = \int_0^{b_n} \left((1 * K_\theta)(t)\right)^2 dt.$$

Since

$$(1 * K_\theta)(t) = \int_0^t K_\theta(s) ds = t^\alpha E_{\alpha,\alpha+1}(\theta t^\alpha),$$

the variance σ_n^2 has the following form

$$\begin{aligned}
\sigma_n^2 &= \int_0^{b_n} t^{2\alpha}(E_{\alpha,\alpha+1}(\theta t^\alpha))^2 dt \\
&= \sum_{n=0}^{+\infty}\left[\frac{\theta^{2n}\int_0^{b_n} t^{2\alpha+2\alpha n}dt}{(\Gamma(\alpha n+\alpha+1))^2} + 2\sum_{k>n}\frac{\theta^{n+k}\int_0^{b_n} t^{2\alpha+\alpha n+\alpha k}dt}{\Gamma(\alpha(n+1)+1)\Gamma(\alpha(k+1)+1)}\right] \\
&= \sum_{n=0}^{+\infty}\left[\frac{\theta^{2n}{b_n}^{2\alpha+2\alpha n+1}}{(\Gamma(\alpha n+\alpha+1))^2(2\alpha+2\alpha n+1)} + 2\sum_{k>n}\frac{\theta^{n+k}{b_n}^{2\alpha+\alpha n+\alpha k+1}}{\Gamma(\alpha(n+1)+1)\Gamma(\alpha(k+1)+1)(2\alpha+\alpha n+\alpha k+1)}\right]
\end{aligned}$$

We know that $2\alpha+\alpha n+\alpha k+1 < 2\alpha+2\alpha(n\vee k)+2$ and $2\alpha+2\alpha n+1 < 2\alpha+2\alpha n+2$, then

$$\begin{aligned}
\sigma_n^2 &\geq \frac{{b_n}^{2\alpha+1}}{2}\sum_{n=0}^{+\infty}\left[\frac{\theta^{2n}{b_n}^{2\alpha n}}{(\Gamma(\alpha n+\alpha+1))^2(\alpha+\alpha n+1)} + 2\sum_{k>n}\frac{\theta^{n+k}{b_n}^{\alpha n+\alpha k}}{\Gamma(\alpha(n+1)+1)\Gamma(\alpha(k+1)+1)(\alpha+\alpha k+1)}\right] \\
&= \frac{{b_n}^{2\alpha+1}}{2}\sum_{n=0}^{+\infty}\left[\frac{\theta^{2n}{b_n}^{2\alpha n}}{\Gamma(\alpha n+\alpha+1)\Gamma(\alpha n+\alpha+2)} + 2\sum_{k>n}\frac{\theta^{n+k}{b_n}^{\alpha n+\alpha k}}{\Gamma(\alpha(n+1)+1)\Gamma(\alpha(k+1)+2)}\right].
\end{aligned}$$

We have that $\Gamma(\alpha n+\alpha+1) < \Gamma(\alpha n+\alpha+2)$ for every n. Hence

$$\begin{aligned}
\sigma_n^2 &\geq \frac{{b_n}^{2\alpha+1}}{2}\sum_{n=0}^{+\infty}\left[\frac{\theta^{2n}{b_n}^{2\alpha n}}{\Gamma(\alpha n+\alpha+2)^2} + 2\sum_{k>n}\frac{\theta^{n+k}{b_n}^{\alpha n+\alpha k}}{\Gamma(\alpha(n+1)+2)\Gamma(\alpha(k+1)+2)}\right] \\
&= \frac{{b_n}^{2\alpha+1}}{2}(E_{\alpha,\alpha+2}(\theta {b_n}^\alpha))^2.
\end{aligned}$$

The conclusion follows from the representations of Mittag-Leffler functions and the expansions for $\mid z\mid$ going to $+\infty$ in (4.12).

4.4 Simulations

Let $n\geq 1$, for $\theta>0$, $a_n=\lfloor n^{\beta+1}\rfloor$ and for $\theta<0$, $a_n=\lfloor \beta n\ln(n)\rfloor$ with β as in Theorem 4.3.1 $t_k=\frac{k}{n}$ for $0\leq k\leq a_n$ and $h_n=1/n$. We simulate 20 trajectories of the Volterra Ornstein-Uhlenbeck process with $X_0=0$ using the factor approximation studied in Section 1.4. The factor approximation process X^N, where N represent the number of factors in the simulation, is obtained at points t_k, $k\in\{0,\dots,a_n\}$. Since $Z_{t_k}=\int_0^{t_k} L(t_k-u)X_u du$, with $L(t)=\frac{t^{-\alpha}}{\Gamma(1-\alpha)}$, we cannot observe directly the process Z. For this reason, to implement the estimator $\bar{\theta}_n(a)$ we use an approximation of the process Z based on the factor-simulation of X. We fix $N=20$ and $\alpha=0.6$. We compute the value of the estimator $\bar{\theta}_n(a)$ in (4.10) for different values of θ and n and present them in the tables below.

n	10000	20000	30000
$\bar{\theta}_n(a)$	2.06	1.99	2.00

Table 4.1: Estimator for the parameter $\theta = 2$

n	30000	50000	100000
$\bar{\theta}_n(a)$	-1.23	-1.96	-2.01

Table 4.2: Estimator for the parameter $\theta = -2$

Bibliography

Eduardo Abi Jaber. Lifting the Heston model. *Quantitative Finance*, 19(12):1995–2013, 2019a.

Eduardo Abi Jaber. The Laplace transform of the integrated Volterra Wishart process. *arXiv preprint arXiv:1911.07719*, 2019b.

Eduardo Abi Jaber and Omar El Euch. Markovian structure of the Volterra Heston model. *Statistics & Probability Letters*, 149:63–72, 2019a.

Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019b.

Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

Eduardo Abi Jaber, Christa Cuchiero, Martin Larsson, and Sergio Pulido. A weak solution theory for stochastic Volterra equations of convolution type. *arXiv preprint arXiv:1909.01166 to appear in Annals of Applied Probability*, 2021a.

Eduardo Abi Jaber, Christa Cuchiero, and Sergio Pulido. Polynomial Volterra processes. *In preparation*, 2021b.

Eduardo Abi Jaber, Enzo Miller, and Huyên Pham. Markowitz portfolio selection for multivariate affine and quadratic Volterra models. *SIAM Journal on Financial Mathematics*, 12(1): 369–409, 2021c.

Eduardo Abi Jaber, Enzo Miller, and Huyên Pham. Linear–quadratic control for a class of stochastic Volterra equations: solvability and approximation. *arXiv preprint arXiv:1911.01900 to appear in The Annals of Applied Probability*, 2021d.

Damien Ackerer and Damir Filipović. Linear credit risk models. *Finance and Stochastics*, 24 (1):169–214, 2020.

Damien Ackerer, Damir Filipović, and Sergio Pulido. The Jacobi stochastic volatility model. *Finance and Stochastics*, 22(3):667–700, 2018.

Aurélien Alfonsi and Ahmed Kebaier. Approximation of stochastic volterra equations with kernels of completely monotone type. *arXiv preprint arXiv:2102.13505*, 2021.

Elisa Alòs, Jorge A León, and Josep Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4):571–589, 2007.

Ishwar V Basawa and David John Scott. *Asymptotic optimal inference for non-ergodic models.* Lecture Notes in Statistics, 17, Springer, 1983.

Nicole Bäuerle and Sascha Desmettre. Portfolio optimization in fractional and rough Heston models. *SIAM Journal on Financial Mathematics*, 11(1):240–273, 2020.

Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

Christian Bayer, Peter K Friz, Archil Gulisashvili, Blanka Horvath, and Benjamin Stemper. Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, 19(5):779–798, 2019.

Christian Bayer, Jinniao Qiu, and Yao Yao. Pricing options under rough volatility with backward SPDEs. *arXiv preprint arXiv:2008.01241*, 2020a.

Christian Bayer, Raúl Tempone, and Sören Wolfers. Pricing american options by exercise rate optimization. *Quantitative Finance*, 20(11):1749–1760, 2020b.

Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20:74, 2019.

Rachid Belfadli, Khalifa Es-Sebaiy, and Youssef Ouknine. Parameter estimation for fractional Ornstein-Uhlenbeck processes: Non-ergodic case. *Frontiers in Science and Engineering*, 1: 1–16, 2011.

Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

Mikkel Bennedsen, Asger Lunde, and and Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 2021. doi: 10.1093/jjfinec/nbaa049.

Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

Jaya PN Bishwal. Minimum contrast estimation in fractional Ornstein-Uhlenbeck process: Continuous and discrete sampling. *Fractional Calculus and Applied Analysis*, 14(3):375–410, 2011.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

Giorgia Callegaro, Martino Grasselli, and Gilles Pagès. Fast hybrid schemes for fractional Riccati equations (rough is not so tough). *Mathematics of Operations Research*, 46(1):221–254, 2021.

Philippe Carmona, Laure Coutin, and Gérard Montseny. Approximation of some Gaussian processes. *Statistical inference for stochastic processes*, 3(1-2):161–171, 2000.

Peter Carr and Dilip Madan. Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

Peggy Cénac and Khalifa Es-Sebaiy. Almost sure central limit theorems for random ratios and applications to LSE for fractional Ornstein–Uhlenbeck processes. *Probability and Mathematical Statistics*, 35(2):285–300, 2015.

Etienne Chevalier, Sergio Pulido, and Elizabeth Zúñiga. American options in the Volterra Heston model. *arXiv preprint arXiv:2103.11734*, 2021.

Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.

Fabienne Comte, Laure Coutin, and Eric Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

John Cox, Jonathan Ingersoll, and Stephen Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 02 1985. doi: 10.2307/1911242.

Christa Cuchiero and Josef Teichmann. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *Journal of Evolution Equations*, pages 1–48, 2020.

Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740, 2012.

Aditi Dandapani, Paul Jusselin, and Mathieu Rosenbaum. From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect. *arXiv preprint arXiv:1907.06151*, 2019.

Adrian A Drăgulescu and Victor M Yakovenko. Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative finance*, 2(6):443–453, 2002.

Darrell Duffie, Damir Filipović, and Walter Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053, 2003.

Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.

Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.

Omar El Euch, Masaaki Fukasawa, and Mathieu Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2018.

Mohamed El Machkouri, Khalifa Es-Sebaiy, and Youssef Ouknine. Least squares estimator for non-ergodic Ornstein–Uhlenbeck processes driven by Gaussian processes. *Journal of the Korean Statistical Society*, 45(3):329–341, 2016.

Khalifa Es-Sebaiy and Djibril Ndiaye. On drift estimation for non-ergodic fractional Ornstein-Uhlenbeck process with discrete observations. *Afrika Statistika*, 9(1):615–625, 2014.

Khalifa Es-Sebaiy, Fares Alazemi, and Mishari Al-Foraih. Least squares type estimation for discretely observed non-ergodic Gaussian Ornstein-Uhlenbeck processes. *Acta Mathematica Scientia*, 39(4):989–1002, 2019.

Omar El Euch, Masaaki Fukasawa, Jim Gatheral, and Mathieu Rosenbaum. Short-term at-the-money asymptotics under stochastic volatility models. *SIAM Journal on Financial Mathematics*, 10(2):491–511, 2019.

Fang Fang and Cornelis W Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2009.

Damir Filipović and Martin Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.

Damir Filipović, Eberhard Mayerhofer, and Paul Schneider. Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111, 2013.

Damir Filipović, Elise Gourier, and Loriano Mancini. Quadratic variance swap models. *Journal of Financial Economics*, 119(1):44–68, 2016.

Damir Filipović, Martin Larsson, and Anders B Trolle. Linear-rational term structure models. *The Journal of Finance*, 72(2):655–704, 2017.

Martin Forde and Hongzhong Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1):114–145, 2017.

Jean-Pierre Fouque and Ruimeng Hu. Optimal portfolio under fractional stochastic environment. *Mathematical Finance*, 29(3):697–734, 2019.

Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

Masaaki Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

Masaaki Fukasawa. Volatility has to be rough. *Quantitative Finance*, 21(1):1–8, 2021.

Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is volatility rough? *arXiv preprint arXiv:1905.04852*, 2019.

Josselin Garnier and Knut Solna. Correction to Black–Scholes formula due to fractional stochastic volatility. *SIAM Journal on Financial Mathematics*, 8(1):560–588, 2017.

Jim Gatheral. The volatility surface: a practitioner's guide. John Wiley & Sons. *Inc., Hoboken, New Jersey*, 2006.

Jim Gatheral and Martin Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23(3):501–533, 2019.

Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

Rudolf Gorenflo, Joulia Loutchko, and Yuri Luchko. Computation of the Mittag-Leffler function and its derivatives. *Fractional Calculus and Applied Analysis*, 5:491–518, 2002.

Ludovic Goudenège, Andrea Molent, and Antonino Zanette. Machine learning for pricing American options in high-dimensional Markovian and non-Markovian models. *Quantitative Finance*, 20(4):573–591, 2020.

Gustaf Gripenberg, Stig-Olof Londen, and Olof Staffans. *Volterra Integral and Functional Equations*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1990. doi: 10.1017/CBO9780511662805.

Hamza Guennoun, Antoine Jacquier, Patrick Roome, and Fangwei Shi. Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics*, 9(3):1017–1045, 2018.

Bingyan Han and Hoi Ying Wong. Merton's portfolio problem under Volterra Heston model. *Finance Research Letters*, 39:101580, 2020a.

Bingyan Han and Hoi Ying Wong. Mean–variance portfolio selection under Volterra Heston model. *Applied Mathematics & Optimization*, pages 1–28, 2020b.

Bingyan Han and Hoi Ying Wong. Time-inconsistency with rough volatility. *arXiv preprint arXiv:1907.11378*, 2020c.

Philipp Harms and David Stefanovits. Affine representations of fractional processes with applications in mathematical finance. *Stochastic Processes and their Applications*, 129(4):1185–1228, 2019.

Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

Steven L Heston and Alberto G Rossi. A spanning series approach to options. *The Review of Asset Pricing Studies*, 7(1):2–42, 2017.

Blanka Horvath, Antoine Jack Jacquier, and Aitor Muguruza. Functional central limit theorems for rough volatility. *SSRN Electronic Journal*, 2017.

Yaozhong Hu and David Nualart. Parameter estimation for fractional Ornstein–Uhlenbeck processes. *Statistics & probability letters*, 80(11-12):1030–1038, 2010.

Yaozhong Hu and Jian Song. Parameter estimation for fractional Ornstein–Uhlenbeck processes with discrete observations. In *Malliavin calculus and stochastic analysis*, pages 427–442. Springer, 2013.

Yaozhong Hu, David Nualart, and Hongjuan Zhou. Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. *Statistical Inference for Stochastic Processes*, 22(1):111–142, 2019.

Thibault Jaisson and Mathieu Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 26(5):2860–2882, 2016.

Martin Keller-Ressel, Martin Larsson, and Sergio Pulido. Affine rough models. *arXiv preprint arXiv:1812.08486. Rough Volatility, Chapter, SIAM, to appear*, 2021.

Marina L Kleptsyna and Alain Le Breton. Statistical analysis of the fractional Ornstein–Uhlenbeck type process. *Statistical Inference for stochastic processes*, 5(3):229–248, 2002.

Kęstutis Kubilius, Yuliya Mishura, Kostiantyn Ralchenko, and Oleg Seleznjev. Consistency of the drift parameter estimator for the discretized fractional Ornstein–Uhlenbeck process with Hurst index $h \in (0, \frac{1}{2})$. *Electronic Journal of Statistics*, 9(2):1799–1825, 2015.

Yury A Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer, 2004.

Damien Lamberton and Gilles Pagès. Sur l'approximation des réduites. In *Annales de l'IHP Probabilités et statistiques*, volume 26, pages 331–355, 1990.

Bernard Lapeyre and Jérôme Lelong. Neural network regression for Bermudan option pricing. *arXiv preprint arXiv:1907.06474*, 2019.

Changpin Li and Chunxing Tao. On the fractional Adams method. *Computers & Mathematics with Applications*, 58(8):1573–1588, 2009.

Robert S Liptser and Albert N Shiryaev. *Statistics of random processes II: Applications.* Second Edition, Applications of Mathematics, Springer, 2001.

Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. Rough volatility: evidence from option prices. *IISE transactions*, 50(9):767–776, 2018.

Francis A Longstaff and Eduardo S Schwartz. Valuing American options by simulation: a simple least-squares approach. *The review of financial studies*, 14(1):113–147, 2001.

Ryan McCrickerd and Mikko S Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11):1877–1886, 2018.

Alexey Anatol'evich Muravlev. Representation of a fractional Brownian motion in terms of an infinite-dimensional Ornstein-Uhlenbeck process. *Russian Mathematical Surveys*, 66(2): 439–441, 2011.

Philip E Protter. Stochastic differential equations. In *Stochastic integration and differential equations*, pages 249–361. Springer, 2005.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

Mehdi Tomas and Mathieu Rosenbaum. From microscopic price dynamics to multidimensional rough volatility models. *arXiv preprint arXiv:1910.13338 to appear in Advances in Applied Probability*, 2021.

Ciprian A Tudor and Frederi G Viens. Statistical aspects of the fractional stochastic calculus. *The Annals of Statistics*, 35(3):1183–1212, 2007.

George E Uhlenbeck and Leonard S Ornstein. On the theory of the Brownian motion. *Physical review*, 36(5):823–841, 1930.

Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

Frederi Viens and Jianfeng Zhang. A martingale approach for fractional Brownian motions and related path dependent PDEs. *The Annals of Applied Probability*, 29(6):3489–3540, 2019.

Hanxiao Wang, Jiongmin Yong, and Jianfeng Zhang. Path dependent Feynman-Kac formula for forward backward stochastic Volterra integral equations. *arXiv preprint arXiv:2004.05825*, 2020.

Weilin Xiao, Weiguo Zhang, and Weidong Xu. Parameter estimation for fractional Ornstein–Uhlenbeck processes at discrete observation. *Applied Mathematical Modelling*, 35(9):4196–4207, 2011.

Haiping Ye, Jianming Gao, and Yongsheng Ding. A generalized Gronwall inequality and its application to a fractional differential equation. *Journal of Mathematical Analysis and Applications*, 328(2):1075–1081, 2007.

ÉCOLE DOCTORALE
de mathématiques
Hadamard (EDMH)

Titre: Processus de Volterra et applications en finance

Mots clés: Processus de Volterra, équations de Riccati-Volterra, volatilité rugueuse, option Américaine, moments de processus de Volterra, estimation paramètrique

Résumé: Cette thèse est consacrée à l'étude des processus de Volterra et leur utilisation en finance. Nous commençons par rappeler certaines propriétés de ces processus que nous utiliserons tout au long de notre travail. La seconde partie porte sur l'étude d'un problème d'arrêt optimal, la valorisation d'une option Américaine dans un modèle de Heston-Volterra. Pour certains choice de paramètres, ce modèle est une version dite rugueuse du bien connu modèle de Heston. Nous nous concentrons sur la convergence des prix dans une suite de modèles de grande dimension, approchant le modèle original, vers les prix dans le modèle limite de Volterra. Dans le troisième chapitre de ce travail, nous étudions les moments de processus polynômiaux de Volterra. Nous proposons des méthodes de calcul des moments de ces processus et montrons qu'ils ont certaines propriétés en commun avec les diffusions polynômiales classiques. Nous concluons ce travail en nous intéressant dans le quatrième chapitre à des problèmes plus statistiques. Nous abordons le problème d'estimation du paramètre de vitesse de retour à la moyenne d'un processus d'Ornstein-Uhlenbeck de Volterra. Nous montrons que nos estimateurs, basés sur des observations continues ou discrètes du processus sont consistants.

Title: Volterra processes and applications in finance

Keywords: Volterra processes, Riccati-Volterra equations, rough volatility, American options, moments of Volterra processes, parametric estimation

Abstract: The present thesis is devoted to the study of stochastic Volterra processes and their applications to finance. We begin by recalling some properties of these processes that will be used throughout this work. The second part focuses on the study of an optimal stopping problem, namely the problem of pricing American options in the Volterra Heston model. For a particular choice of the parameters, this model is a rough version of the widely-known Heston model. We concentrate on the convergence of the prices in an approximating sequence of high dimensional models towards the prices in the limiting Volterra model. The third part of this work is devoted to the study of the moments for polynomial Volterra processes. We develop some methods for computing moments of these processes and show that they share some properties with classical polynomial diffusions. In the last part, we shift our attention to more statistical matters. We tackle the problem of the drift estimation for the Volterra Ornstein-Uhlenbeck process. We prove the consistency of our drift estimators based on continuous and discrete observations of the process.

Maison du doctorat de l'Université Paris-Saclay
$2^{\text{ème}}$ etage aile ouest, Ecole normale supérieure Paris-Saclay
4 avenue des Sciences,
91190 Gif sur Yvette, France