Quantitative Finance under rough volatility

Omar El Euch

THÈSE

présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES DE SORBONNE UNIVERSITÉS

Spécialité : Mathématiques

par

Omar El Euch

Quantitative Finance Under Rough Volatility

Soutenue le 25 Septembre 2018 devant un jury composé de :

Walter Schachermayer	*University of Vienna*	rapporteur
Josef Teichmann	*ETH Zurich*	rapporteur
Bruno Bouchard	*Université Paris-Dauphine*	examinateur
Jean-Philippe Bouchaud	*Capital Fund Management*	examinateur
Peter Friz	*TU Berlin*	examinateur
Jean Jacod	*Université Pierre et Marie Curie*	examinateur
Gilles Pages	*Université Pierre et Marie Curie*	examinateur
Peter Tankov	*ENSAE*	examinateur
Nizar Touzi	*Ecole Polytechnique*	examinateur
Mathieu Rosenbaum	*Ecole Polytechnique*	directeur de thèse

Remerciements

Je tiens en premier lieu à exprimer ma plus profonde gratitude à Mathieu Rosenbaum pour son encadrement scientifique, son soutien et sa confiance. Sous sa direction, j'ai passé trois merveilleuses années pendant lesquelles j'étais totalement épanoui. Je lui suis infiniment reconnaissant!

Je remercie Walter Schachermayer et Josef Teichamnn d'avoir accepté de rapporter ma thèse. Je suis très honoré par leur lecture attentive de ce manuscrit et leur intéret pour mon travail. Je remercie également Bruno Bouchard, Jean-Philippe Bouchaud, Peter Friz, Jean Jacod, Gilles Pagès et Peter Tankov d'examiner ma thèse et de participer à ma soutenance.

Je voudrais particulièrement exprimer ma gratitude à Nizar Touzi pour m'avoir transmis la passion des mathématiques financières et encouragé à faire cette thèse. Je suis honoré d'avoir travaillé avec lui et de l'avoir au sein du jury de ma thèse.

Je remercie également mes collaborateurs Eduardo Abi Jaber, Jim Gatheral, Masaaki Fukasawa, Thibaut Mastrolia et Othmane Mounjid. Ce fut un immense plaisir de travailler avec chacun d'eux!

Toute ma reconnaissance va aussi à Marcos Costa Santos Carreira, Christa Cuchiero, Stefano De Marco, Stefan Gerhold, Julien Guyon, Blanka Horvath, Antoine Jacquier et Alexandre Richard. Les diverses discussions et interactions que j'ai eu avec chacun d'eux ont permis le bon avancement de mon travail.

Je suis reconnaissant à tous mes professeurs de l'Ecole Polytechnique et du Master "Probabilités et Finance", en particulier Emmanuel Gobet, Nicole El Karoui et Aurélien Alfonsi pour m'avoir fait découvrir les mathématiques financières.

Un grand merci à tous les doctorants de l'Ecole Polytechnique et de Jussieu, en particulier Aditi, Eric, Hadrien, Heythem, Paul, Saad et surtout Pamela pour leur bonne humeur et pour avoir rendu mon quotidien encore plus agréable durant ces trois dernières années.

Je remercie égalemement tout le secrétariat du CMAP et du LPMA et plus particuliérement Nasséra Naar pour sa disponibilité et son aide pendant trois ans.

Pour finir, je remercie mes parents, ma soeur, mes cousines Chema et Hend, ma famille, mes amis et Mariem pour leurs encouragements et leur soutien durant cette thèse et bien avant.

Résumé

Cette thèse a pour objectif la compréhension de plusieurs aspects du caractère rugueux de la volatilité observé de manière universelle sur les actifs financiers. Ceci est fait en six étapes.

Dans une première partie, on explique cette propriété à partir des comportements typiques des agents sur le marché. Plus précisément, on construit un modèle de prix microscopique basé sur les processus de Hawkes reproduisant les faits stylisés importants de la microstructure des marchés. En étudiant le comportement du prix à long terme, on montre l'émergence d'une version rugueuse du modèle de Heston (appelé modèle rough Heston) avec effet de levier.

En utilisant ce lien original entre les processus de Hawkes et les modèles de Heston, on calcule dans la deuxième partie de cette thèse la fonction caractéristique du log-prix du modèle rough Heston. Cette fonction caractéristique est donnée en terme d'une solution d'une équation de Riccati dans le cas du modèle de Heston classique. On montre la validité d'une formule similaire dans le cas du modèle rough Heston, où l'équation de Riccati est remplacée par sa version fractionnaire. Cette formule nous permet de surmonter les difficultés techniques dues au caractère non markovien du modèle afin de valoriser des produits dérivés.

Dans la troisième partie, on aborde la question de la gestion des risques des produits dérivés dans le modèle rough Heston. On présente des stratégies de couverture utilisant comme instruments l'actif sous-jacent et la courbe variance forward. Ceci est fait en spécifiant la structure markovienne infini-dimensionnelle du modèle.

Étant capable de valoriser et couvrir les produits dérivés dans le modèle rough Heston, nous confrontons ce modèle à la réalité des marchés financiers dans la quatrième partie. Plus précisément, on montre qu'il reproduit le comportement de la volatilité implicite et historique. On montre également qu'il génère l'effet Zumbach qui est une asymétrie par inversion du temps observée empiriquement sur les données financières.

On étudie dans la cinquième partie le comportement limite de la volatilité implicite à la monnaie à faible maturité dans le cadre d'un modèle à volatilité stochastique général (incluant le modèle rough Bergomi), en appliquant un développement de la densité du prix de l'actif.

Alors que l'approximation basée sur les processus de Hawkes a permis de traiter plusieurs questions relatives au modèle rough Heston, nous examinons dans la sixième partie une approximation markovienne s'appliquant sur une classe plus générale de modèles à volatilité rugueuse. En utilisant cette approximation dans le cas particulier du modèle rough Heston, on obtient une méthode numérique pour résoudre les équations de Riccati fractionnaires.

Enfin, nous terminons cette thèse en étudiant un problème non lié à la littérature sur la volatilité rugueuse. Nous considérons le cas d'une plateforme cherchant le meilleur système de make-take fees pour attirer de la liquidité. En utilisant le cadre principal-agent, on décrit le meilleur contrat à proposer au market maker ainsi que les cotations optimales affichées par ce dernier. Nous montrons également que cette politique conduit à une meilleure liquidité et à une baisse des coûts de transaction pour les investisseurs.

Abstract

The aim of this thesis is to study various aspects of the rough behavior of the volatility observed universally on financial assets. This is done in six steps.

In the first part, we investigate how rough volatility can naturally emerge from typical behaviors of market participants. To do so, we build a microscopic price model based on Hawkes processes in which we encode the main features of the market microstructure. By studying the asymptotic behavior of the price on the long run, we obtain a rough version of the Heston model exhibiting rough volatility and leverage effect.

Using this original link between Hawkes processes and the Heston framework, we compute in the second part of the thesis the characteristic function of the log-price in the rough Heston model. In the classical Heston model, the characteristic function is expressed in terms of a solution of a Riccati equation. We show that rough Heston models enjoy a similar formula, the Riccati equation being replaced by its fractional version. This formula enables us to overcome the non-Markovian nature of the model in order to deal with derivatives pricing.

In the third part, we tackle the issue of managing derivatives risks under the rough Heston model. We establish explicit hedging strategies using as instruments the underlying asset and the forward variance curve. This is done by specifying the infinite-dimensional Markovian structure of the rough Heston model.

Being able to price and hedge derivatives in the rough Heston model, we challenge the model to practice in the fourth part. More precisely, we show the excellent fit of the model to historical and implied volatilities. We also show that the model reproduces the Zumbach's effect, that is a time reversal asymmetry which is observed empirically on financial data.

While the Hawkes approximation enabled us to solve the pricing and hedging issues under the rough Heston model, this approach cannot be extended to an arbitrary rough volatility model. We study in the fifth part the behavior of the at-the-money implied volatility for small maturity under general stochastic volatility models.

In the same spirit as the Hawkes approximation, we look in the sixth part of this thesis for a tractable Markovian approximation that holds for a general class of rough volatility models. By applying this approximation on the specific case of the rough Heston model, we derive a numerical scheme for solving fractional Riccati equations.

Finally, we end this thesis by studying a problem unrelated to rough volatility. We consider an exchange looking for the best make-take fees system to attract liquidity in its platform. Using a principal-agent framework, we describe the best contract that the exchange should propose to the market maker and provide the optimal quotes displayed by the latter. We also argue that this policy leads to higher quality of liquidity and lower trading costs for investors.

Keywords

Market microstructure, high frequency trading, leverage effect, rough volatility, Hawkes processes, limit theorems, Heston model, rough Heston model, fractional Brownian motion, fractional Riccati equation, forward variance curve, affine Volterra processes, stochastic Volterra equations, Markovian representation, stochastic invariance, Zumbach's effect, Edgeworth expansion, rough volatility models, make-take fees, market making, financial regulation, principal-agent problem, stochastic control.

List of papers being part of this thesis

- O. El Euch, M. Fukasawa and M. Rosenbaum, *The microstructural foundations of leverage effect and rough volatility*, Finance and Stochastics, 22(2), 241-280, 2018.
- O. El Euch and M. Rosenbaum, *The characteristic function of rough Heston models*, to appear in Mathematical Finance.
- O. El Euch and M. Rosenbaum, *Perfect hedging in rough Heston models*, to appear in Annals of Applied Probability.
- E. Abi Jaber and O. El Euch, *Markovian structure of the Volterra Heston model*, submitted, 2018.
- O. El Euch, J. Gatheral and M. Rosenbaum, *Roughening Heston*, submitted, 2018.
- O. El Euch, J. Gatheral, R. Radoicic and M. Rosenbaum, *The Zumbach effect under rough Heston*, submitted, 2018.
- O. El Euch, M. Fukasawa, J. Gatheral and M.Rosenbaum, *Short-term at-the-money asymptotics under stochastic volatility models*, submitted, 2017.
- E. Abi Jaber and O. El Euch, *Multi-factor approximations of rough volatility models*, submitted, 2018.
- O. El Euch, T. Mastrolia, M. Rosenbaum and N. Touzi, *Optimal make-take fees for market making regulation*, submitted, 2018.

Contents

Contents **x**

Introduction **1**

Motivations . . . 1
Outline . . . 3
1 Part I: The microstructural foundations of leverage effect and rough volatility. 4
2 Part II: Characteristic function of rough Heston models . . . 8
3 Part III: Hedging under the rough Heston model . . . 12
4 Part IV: The rough Heston model in practice . . . 17
5 Part V: Short-term behavior of the at-the-money implied volatility under rough volatility . . . 18
6 Part VI: Markovian approximation of rough volatility models . . . 21
7 Part VII: Optimal make-take fees for market making regulation . . . 25

Part I The microstructural foundations of rough volatility **31**

I Microstructural foundations of leverage effect and rough volatility **33**

1 Introduction . . . 33
2 From high frequency features to leverage effect . . . 37
3 From high frequency features to rough volatility . . . 47
4 Proofs . . . 52
I.A Technical results . . . 66
I.B Fractional integrals and derivatives . . . 67
I.C Mittag-Leffler functions . . . 67

Part II Pricing under the rough Heston model **69**

II Characteristic function of rough Heston models **71**

1 Introduction . . . 71
2 From Hawkes processes to rough Heston models . . . 75
3 The characteristic function of multivariate Hawkes processes . . . 81

4 The characteristic function of rough Heston models 84
5 Numerical application 86
6 Proofs 90
II.A Mittag-Leffler functions 103
II.B Wiener-Hopf equations 104
II.C Fractional differential equations 104

Part III Hedging under the rough Heston model **107**

III Perfect hedging in rough Heston models **109**
1 Introduction 109
2 Conditional laws of rough Heston models 112
3 Characteristic function of generalized rough Heston models 113
4 Hedging under generalized rough Heston models 122
5 Proofs 125
III.A Fractional calculus 138
III.B Martingale property of the price in the generalized rough Heston model . . . 142
III.C Moments properties for Hawkes processes 143

IV Markovian structure of the Volterra Heston model **147**
1 Introduction 147
2 Existence and uniqueness of the Volterra Heston model 149
3 Markovian structure 152
4 Application: square-root process in Banach space 156
IV.A Existence results for stochastic Volterra equations 159
IV.B Reminder on stochastic convolutions and resolvents 160

Part IV The rough Heston model in practice **163**

V Roughening Heston **165**
1 Introduction 165
2 The rough Heston model 166
3 Pricing and hedging 166
4 Calibration of the rough Heston model 167
5 A poor man's rough Heston model 170
V.A Numerical solution of the fractional Riccati equation 172
V.B Call option prices using Fourier techniques 174

VI Zumbach's effect in rough Heston model **175**
1 Introduction 175
2 Empirical estimation of the Zumbach effect 176
3 The rough Heston model 177

4 Numerical results 180
VI.A List of indices in the Oxford-Man Institute of Quantitative Finance Realized Library 181
VI.B Proof of (5) 183
VI.C The Zumbach effect in terms of correlations in the stationary regime 183
VI.D Variance computations 184

Part V Short-term behavior of the at-the-money implied volatility under rough volatility models 187

VII Short-term at-the-money asymptotics under stochastic volatility models 189
1 Introduction 189
2 Framework 190
3 Asymptotic expansions 193
4 Asymptotics for at-the-money skew and curvature 199
5 The rough Bergomi model 200
VII.A Conditional expectations of Wiener-Itô integrals 205

Part VI Markovian approximation of rough volatility models 207

VIII Multi-factor approximation of rough volatility models 209
1 Introduction 209
2 A definition of rough volatility models 212
3 Multi-factor approximation of rough volatility models 213
4 The particular case of the rough Heston model 219
5 Proofs 225
VIII.A Stochastic convolutions and resolvents 237
VIII.B Some existence results for stochastic Volterra equations 240
VIII.C Linear Volterra equation with continuous coefficients 244

Part VII Optimal make-take fees for market making regulation 249

IX Optimal make-take fees for market making regulation 251
1 Introduction 251
2 The model 253
3 Solving the market maker's problem 256
4 Designing the optimal contract 258
5 Impact of the presence of the exchange on market quality and comparison with [AS08, GLFT13] 264
IX.A Predictable representation 271
IX.B Dynamic programming principle and contract representation 272

IX.C Exchange Hamiltonian maximization 276
IX.D On the verification argument for the exchange problem 277

Bibliography **279**

Introduction

The aim of this work is to study various aspects of rough volatility models such as the microstructural foundations of rough volatility, the pricing and hedging under these models and finding relevant model approximations for numerical computations. We also discuss an optimal make-take fees policy for market making regulation.

Motivations

The rough nature of the volatility is nowadays considered a universal feature of financial data. It has been first discovered in [GJR18] and then confirmed in [BLP16]. Indeed, empirical studies over a very wide range of assets of volatility time series have shown that the volatility exhibits a dynamic that is rougher than that of a Brownian motion. Moreover, including this stylized fact into financial models leads to a better fit of the volatility surface, see [BFG16, Fuk17]. The first step of this thesis is to understand how such feature can be generated by answering the following question:

Question 1. *Can the rough behavior of the volatility be explained from microscopic interactions between agents?*

In order to answer this question, we build a tick by tick price model encoding the microscopic stylized facts observed in the market in the context of high frequency trading. This model is based on Hawkes processes. By studying the asymptotic behavior of our microscopic price dynamic in the long run, we obtain in the limit a rough version of the Heston model.

Since the discovery of this empirical feature, many works aim at rethinking classical stochastic volatility models in order to account for the rough behavior of the volatility. A way to do so is by replacing the classical Brownian motion by a fractional one with a small Hurst parameter around 0.1. However, due to the non-Markovian nature of the fractional Brownian motion, difficulties are encountered in practice when it comes to derivatives pricing. Under the classical Heston model, efficient numerical methods for option pricing have been developed in [AMST07, CM99, KJ05, Lew01] by using the explicit formula for the characteristic function of the asset log-price established in [Hes93]. The convergence result linking Hawkes processes to rough volatility may help us to extend this formula to the rough case and to answer the following question:

Question 2. *Can we obtain a formula of the characteristic function in the rough Heston model?*

By answering the question above, numerical methods for option pricing can be developed. However, in practice, a pricing procedure is not enough. We should also be able to manage derivatives risks. This makes the question bellow important so that such model can be used in the industry:

Question 3. *How do we hedge derivatives in the rough Heston model?*

Once we are able to price and hedge derivatives in the rough Heston model, it will be natural to challenge the model to practice. By introducing the rough volatility feature into the Heston framework, we expect a better fit of the model to the historical and implied volatility. Hence, we raise the following issue:

Question 4. *How does the rough Heston model behave in practice?*

While the Hawkes approximation will enable us to answer the questions above, this approach is only specific to the rough Heston case and cannot be extended to an arbitrary rough volatility model. We can use a direct approach by applying small-time expansions of the density of the asset price to answer our next question:

Question 5. *How does the at-the-money implied volatility of rough volatility models behave in the short term?*

In the same spirit as the Hawkes approximation, it is also natural to look for a tractable Markovian approximation that can be used to tackle derivatives pricing and hedging as well as numerical simulation of a general class of rough volatility models. This leads to the following question:

Question 6. *Can we approximate a rough volatility model by a tractable Markovian one?*

Finally, we end the thesis with a topic unrelated to rough volatility. Indeed, we study the problem of optimal market making regulation. Nowadays, with the fragmentation of financial markets, exchanges are in competition and look for the best way to attract liquidity in their platforms. Hence, they apply a make-take fees system where they subsidize liquidity providers and tax liquidity consumers. Such system led to the emergence of a new type of market makers, such as high frequency traders, that aim at collecting fee rebates. However many studies have shown that these market makers tend to leave the market in a period of stress, see [MSLR17]. Hence, they stop being liquidity providers during a period of time the market needs the most liquidity. It is therefore natural to look for a way to avoid this kind of situation by answering the following question:

Question 7. *What is the optimal make-take fees system that the exchange should apply on its platform?*

Outline

Each question presented above corresponds to a part of the thesis.

In Part I, we answer Question 1 by building a Hawkes-based microscopic price model able to reproduce the main features of market microstructure: high endogoneity of the market, no-arbitrage property, buying/selling asymmetry of liquidity and presence of metaorders. We prove in Chapter I that when the first three of these stylized facts are considered, the microscopic price behaves in the long run as a Heston stochastic volatility model exhibiting leverage effect. When we add the last property (presence of metaroders), we obtain in the limit a rough version of the Heston model exhibiting both leverage effect and rough volatility. Therefore we show that rough volatility is generated from the high endogeneity of the market together with the metaorders splitting phenomenon. Furthermore, we obtain that leverage effect can be at least partially explained from microstructural features.

We answer Question 2 in Part II by using the link between Hawkes processes and the rough Heston model of Part I. While the characteristic function of the log-price under the classical Heston model is given explicitly in terms of the solution of a Riccati equation, we show in Chapter II that this formula can be extended to the rough Heston model, the Riccati equation being replaced by a fractional one. In practice, fractional Riccati equations cannot be solved explicitely, but numerical methods such as the Adams scheme, presented in Chapter II, can be used.

In Part III, we tackle Question 3 by identifying the conditional law of the rough Heston model and studying its infinite-dimensional Markovian structure. In particular, we obtain in Chapter III explicit hedging strategies of derivatives by using as instruments the state variables of the model, namely the underlying asset and the forward variance curve. We also look for a general state space in which the forward variance curve lies in Chapter IV.

In Part IV, we treat Question 4 and examine the fit of the rough Heston model to the market. More precisely, we summarize in Chapter V the results obtained in Parts II and III and show the amazing fit of the rough Heston model to the SPX volatility surface. Then in Chapter VI, we answer a question raised by Jean Philippe Bouchaud about the ability of the model to reproduce a time reversal asymmetry observed in financial time series.

In Part V, we answer Question 5 by applying an Edgeworth small-time expansion of the asset price density under a general stochastic volatility framework. In particular, we obtain the behavior of the at-the-money implied skew and curvature for small maturity and apply these results to the particular case of the rough Bergomi model, see Chapter VII.

Answers to Question 6 can be found in Part VI where we study in Chapter VIII the convergence of multi-factor stochastic volatility models sequence to a general class of rough volatility models. This multi-factor approximation is naturally obtained by smoothing the fractional kernel appearing in the dynamics of the variance process. By applying this approximation on the specific case of the rough Heston model, we derive a numerical scheme for solving fractional Riccati equations appearing in the characteristic function formula.

Finally, Question 7 is studied in Part VII. We consider a framework where one market maker quotes bid and ask prices on a platform held by an exchange who earns transaction costs from each market order. The intensity of market orders arrival is a deterministic function of the spread. In order to attract liquidity in its platform, we provide in Chapter IX the optimal contract that the exchange should propose to the market maker. This is done by using a principal-agent framework: First we compute the market maker optimal quotes for any contract proposed by the exchange. Then knowing the optimal quotes of the market maker, we compute the best contract that the exchange should choose. We finally study the effect of the optimal policy on the liquidity of the market, the profit and loss of the market maker and the exchange and the transaction costs for investors.

Let us now rapidly review the main results of this thesis.

1 Part I: The microstructural foundations of leverage effect and rough volatility.

A very wide range of assets exhibits a rough dynamic of the volatility with the same order of magnitude for the roughness parameter around 0.1 in the Hölder sense. Understanding how such a stylized fact can be generated is of course a natural question. Moreover, it will lead us to rough volatility models that we can manage in practice.

In Chapter I, we aim at understanding how microscopic features of the market at the high frequency scale can give rise in the long run to the crucial macroscopic stylized facts: leverage effect and rough volatility. More precisely, we build a tick by tick price model based on Hawkes processes. Then, we encode in this model the main features of market microstructure and study its long run behavior.

1.1 Building the microscopic model

Inspired by [BDHM13a, JR15], we model the tick by tick price as follows:

$$P_t = N_t^+ - N_t^-,$$

where $N_t = (N_t^+, N_t^-)$ is a bi-dimensional Hawkes process with intensity $\lambda_t = (\lambda_t^+, \lambda_t^-)$ defined by

$$\begin{pmatrix}\lambda_t^+\\ \lambda_t^-\end{pmatrix} = \begin{pmatrix}\mu^+\\ \mu^-\end{pmatrix} + \int_0^t \begin{pmatrix}\varphi_1(t-s) & \varphi_3(t-s)\\ \varphi_2(t-s) & \varphi_4(t-s)\end{pmatrix}.\begin{pmatrix}dN_s^+\\ dN_s^-\end{pmatrix},$$

where μ^+ and μ^- are positive constants and

$$\phi = \begin{pmatrix}\varphi_1 & \varphi_3\\ \varphi_2 & \varphi_4\end{pmatrix} : \mathbb{R}_+ \to \mathcal{M}^2(\mathbb{R}_+).$$

In this framework, $\lambda_t^+ dt$ corresponds to the probability of an upward jump of P between times t and $t+dt$. This probability has three components:

- $\mu_+ dt$, the Poissonian part of the intensity, is the probability that the price goes up because of some exogenous reason.
- $\big(\int_0^t \varphi_1(t-s)dN_s^+\big)dt$, is the probability of upward jump induced by past upward jumps.
- $\big(\int_0^t \varphi_3(t-s)dN_s^-\big)dt$, is the probability of upward jump induced by past downward jumps.

Similar analysis can be made on $\lambda_t^- dt$.

We now present the relevant stylized facts of market microstructure in our setting:

- **Encoding the absence of statistical arbitrage**: At the high-frequency scale, designing a strategy that is profitable on average is a very intricate task, see [ALR14]. We model this fact by assuming that the number of future upward jumps is on average equal to the number of downward jumps, namely

  $$\mathbb{E}[N_t^+] = \mathbb{E}[N_t^-].$$

  This condition is satisfied by taking

  $$\mu^+ = \mu^-, \quad \varphi_1 + \varphi_3 = \varphi_2 + \varphi_4. \tag{1}$$

- **Dealing with the asymmetry of liquidity between the bid and ask side of the order book**: The fact that the ask side of the order book is more liquid than the bid side is a stylized fact commonly observed in the market, see [BCST12, BP09, HS06, HS81, TT12]. In the Hawkes framework, this is translated by an asymmetry in the kernel matrix ϕ such that

  $$\varphi_3 = \beta\varphi_2, \quad \beta > 1. \tag{2}$$

- **The high degree of endogeneity of the market**: Markets are highly endogenous. This means that most of the orders have no real economic motivation but are rather sent in reaction to past orders, see [FS15, HBB13]. Thanks to a population interpretation of Hawkes processes, we can show that the stability condition

  $$\mathscr{S}\Big(\int_0^\infty \phi(s)ds\Big) = \|\varphi_1\|_1 + \beta\|\varphi_2\|_1 < 1,$$

  where $\mathscr{S}$ denotes the spectral radius operator, should almost be violated in order to take into account the high endogeneity of the market. Hence, we assume that this spectral radius is smaller but close to unity. To do so, we index the Hawkes process by a parameter $T>0$ meant to go to infinity. More precisely, we assume that the microscopic price P^T is given by

$$P_t^T = N_t^{T,+} - N_t^{T,-}, \tag{3}$$

where $N_t^T = (N_t^{T,+}, N_t^{T,-})$ is a bi-dimensional Hawkes process indexed by T, with intensity $\lambda_t^T = (\lambda_t^{T,+}, \lambda_t^{T,-})$ defined by

$$\begin{pmatrix}\lambda_t^{T,+}\\ \lambda_t^{T,-}\end{pmatrix} = \begin{pmatrix}\mu^T\\ \mu^T\end{pmatrix} + \int_0^t \phi^T(t-s).\begin{pmatrix}dN_s^{T,+}\\ dN_s^{T,-}\end{pmatrix},$$

with μ^T a positive constant. The kernel matrix $\phi^T : \mathbb{R}_+ \to \mathscr{M}^2(\mathbb{R}_+)$ is given by

$$\phi^T = a_T\phi = a_T\begin{pmatrix}\varphi_1 & \varphi_3\\ \varphi_2 & \varphi_4\end{pmatrix}, \quad a_T \in (0,1),$$

and should satisfy

$$\mathscr{S}(\int_0^\infty \phi(s)ds) = 1, \quad a_T \underset{T\to\infty}{\longrightarrow} 1, \tag{4}$$

so that the high endogeneity of the market is obtained when T is large enough.

1.1.1 Final microscopic model

The final microscopic model for the price P^T is given by (3). In order to encode the stylized facts discussed above, Conditions (1), (2) and (4) shall be satisfied. These conditions impose the following specific structure of the kernel matrix ϕ^T.

Assumption 1. *We assume that*

$$\phi^T = a_T\phi = a_T\begin{pmatrix}\varphi_1 & \beta\varphi_2\\ \varphi_2 & (\beta-1)\varphi_2+\varphi_1\end{pmatrix},$$

with $\varphi_1, \varphi_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that $\|\varphi_1\|_1 + \beta\|\varphi_2\|_1 = 1$, and $a_T \in (0,1)$ converging to unity as T goes to infinity.

1.2 Generating leverage effect

In order to obtain a first non-degenerate limit in the long run, we use the following additional assumption on the asymptotic framework and the kernel matrix.

Assumption 2. *There exist positive parameters λ, μ and m such that*

$$T(1-a_T) \underset{T\to\infty}{\to} \lambda, \quad \mu_T = \mu,$$

and

$$\mathscr{S}(\int_0^\infty x\phi(x)dx) = m < \infty.$$

After a suitable scaling in time and space, we show that the microscopic price converges to a Heston dynamics exhibiting a negative correlation between price and volatility moves.

Result 1. *Under Assumptions 1 and 2, as T tends to infinity, the rescaled microscopic price*

$$\frac{1}{T}P^T_{tT} = \frac{N^{T,+}_{tT} - N^{T,-}_{tT}}{T}, \quad t \in [0,1],$$

converges in law for the Skorokhod topology to the following Heston model:

$$P_t = \frac{1}{1-(\|\varphi_1\|_1 - \|\varphi_2\|_1)}\sqrt{\frac{2}{1+\beta}}\int_0^t \sqrt{X_s}dW_s,$$

with

$$dX_t = \frac{\lambda}{m}\big((\beta+1)\frac{\mu}{\lambda} - X_t\big)dt + \frac{1}{m}\sqrt{\frac{1+\beta^2}{1+\beta}}\sqrt{X_t}dB_t, \quad X_0 = 0,$$

where (W,B) is a correlated bi-dimensional Brownian motion with

$$d\langle W,B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

Result 1 shows that when $\beta > 1$, we obtain leverage effect generated by the asymmetry of liquidity between the bid and ask side of the order book. Note that Conditions (1) and (4) are also important to obtain a non-degenerate macroscopic price limit. To our knowledge, this result is the first in the literature relating in a non ad-hoc way the leverage effect to high frequency dynamics.

1.3 Generating rough volatility

The microscopic price above does not take into account an important stylized fact: The wide presence of metaorders in the market, see [AC01, LL13]. This property is a crucial feature of the market microstructure. It is translated in the Hawkes framework by considering a kernel matrix ϕ^T exhibiting a heavy tail as explained in [JR16b]. This leads us to replace Assumption 2 by the following one.

Assumption 3. *There exist $\alpha \in (1/2,1)$ and $C > 0$ such that*

$$\alpha x^\alpha \int_x^\infty \varphi_1(s) + \beta\varphi_2(s)ds \underset{x\to\infty}{\to} C.$$

Moreover, for some $\lambda^* > 0$ and $\mu > 0$,

$$T^\alpha(1-a_T) \underset{T\to\infty}{\to} \lambda^* > 0, \quad T^{1-\alpha}\mu_T \underset{T\to\infty}{\to} \mu.$$

By incorporating the effect of metaorders into our microscopic model, we obtain that the limiting behavior of the price is different from that in Result 1.

Result 2. *Let* $\lambda = \alpha\lambda^* / \big(C\Gamma(1-\alpha)\big)$. *Under Assumptions 1 and 3, as* T *tends to infinity, the rescaled microscopic price*

$$\sqrt{\frac{1-a_T}{\mu T^\alpha}} P^T_{tT} = \sqrt{\frac{1-a_T}{\mu T^\alpha}} (N^{T,+}_{tT} - N^{T,-}_{tT}), \quad t \in [0,1],$$

converges in the sense of finite dimensional laws to the following rough Heston model:

$$P_t = \frac{1}{1-(\|\varphi_1\|_1 - \|\varphi_2\|_1)} \sqrt{\frac{2}{\beta+1}} \int_0^t \sqrt{Y_s} dW_s,$$

with Y *the unique solution of*

$$Y_t = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda\big((1+\beta) - Y_s\big) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda \sqrt{\frac{1+\beta^2}{\lambda^* \mu(1+\beta)}} \sqrt{Y_s} dB_s,$$

where (W,B) *is a correlated bi-dimensional Brownian motion with*

$$d\langle W,B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}} dt.$$

Furthermore, the process Y_t *has Hölder regularity* $\alpha - 1/2 - \varepsilon$ *for any* $\varepsilon > 0$.

Result 2 shows that the limiting behavior of the microscopic price is still of Heston-type. However, a new fractional kernel $(t-s)^{\alpha-1}$ appears in the limiting dynamics and creates a rough behavior of the volatility. Actually this kernel appears also in the Mandelbrot-van Ness representation of the fractional Brownian motions W^H

$$W^H_t = \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^0 \big((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\big) dW_s + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s.$$

Hence, the tail exponent α is linked to the Hurst parameter of the limiting model $H = \alpha - 1/2$. In particular, this shows that the rough behavior of the volatility is explained by the high degree of endogeneity of the market together with the wide presence of metaorders. Note that even more fundamental arbitrage-based foundations of rough volatility have been recently developed in [JR18].

2 Part II: Characteristic function of rough Heston models

By taking into account the main features of market microstructure, Result 2 shows that after a suitable scaling in time and space, the microscopic price (3) converges in the long run to a rough version of the Heston model. In Chapter II, we use this convergence result to obtain the characteristic function of the rough Heston model. This is done by computing the characteristic function of the microscopic price and then passing to the limit. From this characteristic function formula, efficient numerical methods can be applied in order to price derivatives, see [AMST07, CM99, KJ05, Lew01].

2.1 Link between the rough Heston model and Hawkes processes

In the spirit of Chapter I, we define the rough Heston model as a rough generalization of the celebrated Heston model, where the fractional kernel $(t-s)^{\alpha-1}$ is added in the definition of the variance process:

$$dS_t = S_t\sqrt{V_t}dW_t$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta - V_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\nu\sqrt{V_s}dB_s. \tag{5}$$

The parameters λ, θ, V_0 and ν in (5) are positive and W and B are two Brownian motions with correlation ρ. These parameters play the same role as in the classical Heston model. The additional parameter $\alpha \in (1/2, 1)$ governs the smoothness of the volatility sample paths. The well-definition of this model is established by showing the weak existence and uniqueness of a non-negative solution V of the fractional stochastic differential equation appearing in (5), see Proposition 3 in Chapter II. Moreover, we show that the process V is $\alpha - 1/2 - \varepsilon$ Hölder continuous for any $\varepsilon > 0$. Using the link between Hawkes processes and the Heston framework, we aim at computing the characteristic function of the log-price

$$X_t = \log(S_t/S_0) = \int_0^t \sqrt{V_s}dW_s - \frac{1}{2}\int_0^t V_s ds.$$

We consider the microscopic price P^T defined in (3). We make the following choice of the kernel matrix ϕ^T so that Assumptions 1 and 3 are met.

Assumption 4. *There exists $\beta \geq 0$ such that*

$$a_T = 1 - \lambda T^{-\alpha}, \quad \phi^T = \varphi^T \chi,$$

where

$$\chi = \frac{1}{\beta+1}\begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix}, \quad \varphi^T = a_T\varphi, \quad \varphi = f^{\alpha,1},$$

with $f^{\alpha,1}$ the Mittag-Leffler density function defined in Section II.A of Chapter II.

Result 2 needs to be adapted so that the limiting process is the log-price X_t:

- Result 2 states that $\sqrt{\frac{\theta(1-a_T)}{2\mu T^\alpha}}P_t^T$ converges to the martingale part of X_t, namely $\int_0^t \sqrt{V_s}dW_s$. So that the drift part of X_t appears in the limit, we consider the following modification of the microscopic price:

$$X_t^T = \sqrt{\frac{\theta\lambda}{2\mu T^{2\alpha}}}P_t^T - \frac{\theta\lambda}{2\mu T^{2\alpha}}N_t^{T,+}.$$

- Result 2 leads to a vanishing initial variance in the long run. In order to avoid this phenomenon, we need to change the Poissonian part of the intensity μ_T into a time dependent one denoted by $\hat{\mu}_T(t)$ satisfying the following assumption.

Assumption 5. *The baseline intensity $\hat{\mu}_T$ is given by*

$$\hat{\mu}_T(t) = \mu_T + \xi\mu_T\Big(\frac{1}{1-a_T}(1-\int_0^t \varphi^T(s)ds) - \int_0^t \varphi^T(s)ds\Big),$$

with $\xi > 0$ and $\mu_T = \mu T^{\alpha-1}$ for some $\mu > 0$.

Under these new conditions, we obtain the following result.

Result 3. *Under Assumptions 4 and 5, the sequence of processes $(X_t^T)_{t\in[0,1]}$ converges in law for the Skorokhod topology to $(X_t)_{t\in[0,1]}$ with*

$$V_0 = \theta\xi, \quad \nu = \sqrt{\frac{\lambda\theta(1+\beta^2)}{\mu(1+\beta)^2}}, \quad \rho = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}.$$

2.2 The characteristic function of Hawkes processes

In order to compute the characteristic function of the log-price X_t, we compute the characteristic function of X_t^T and then we pass to the limit. X_t^T being a linear combination of $N_{tT}^{T,+}$ and $N_{tT}^{T,-}$, we need to be able to compute the characteristic function of a multivariate Hawkes process.

Let us consider now a d-dimensional Hawkes process $N = (N^1, \dots, N^d)$ with intensity

$$\lambda_t = \begin{pmatrix} \lambda_t^1 \\ \vdots \\ \lambda_t^d \end{pmatrix} = \mu(t) + \int_0^t \phi(t-s).dN_s,$$

where $\mu : \mathbb{R}_+ \to \mathbb{R}_+^d$ is locally integrable and $\phi : \mathbb{R}_+ \to \mathcal{M}^{\mathbf{d}}(\mathbb{R}_+)$ has integrable components such that

$$\mathcal{S}\Big(\int_0^\infty \phi(s)ds\Big) < 1.$$

We give a population interpretation of this process, as in [HO74], in which we consider d types of individuals and for for each type, an individual can be either a migrant or the descendant of a migrant such that

- Migrants of type $k \in \{1,..,d\}$ arrive as a non-homogenous Poisson process with rate $\mu_k(t)$.
- Each migrant of type $k \in \{1,..,d\}$ gives birth to children of type $j \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{j,k}(t)$.
- Each child of type $k \in \{1,..,d\}$ also gives birth to other children of type $j \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{j,k}(t)$.

Then, for $k \in \{1,..,d\}$, N_t^k can be taken as the number up to time t of migrants and children born with type k. Using this population interpretation, we are able to characterize the law the Hawkes process and compute its characterisitic function.

Result 4. *For any $a \in \mathbb{R}^d$*

$$\mathbb{E}[\exp(ia.N_t)] = \exp\Big(\int_0^t \big(C(a,t-s) - \mathbf{1}\big).\mu(s)ds\Big),$$

where $C : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{C}^d$ is solution of the following integral equation:

$$C(a,t) = \exp\big(ia + \int_0^t \phi^*(s).(C(a,t-s) - \mathbf{1})ds\big),$$

with $\phi^*(s)$ the transpose of $\phi(s)$.

2.3 The characteristic function of rough Heston models

From Result 4, we are able to compute the characteristic function of X_t^T

$$L^T(a,t) = \mathbb{E}[\exp(iaX_t^T)], \quad a \in \mathbb{R}.$$

Moreover as a consequence of Result 3, $L^T(a,t)$ converges to the characteristic function of the log price X_t

$$L(a,t) = \mathbb{E}[\exp(iaX_t)], \quad a \in \mathbb{R},$$

as T goes large. This enables us to obtain a formula for $L(a,t)$. Let $I^{1-\alpha}$ be the fractional integral operator defined by

$$I^{1-\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} f(s)ds,$$

and D^α be the fractional derivative operator defined by

$$D^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t (t-s)^{-\alpha} f(s)ds.$$

We have the following result.

Result 5. *Assume that $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$. Then, the characteristic function of the log price of the rough Heston model is given by*

$$L(a,t) = \exp\Big(\int_0^t (\theta\lambda + V_0 \frac{s^{-\alpha}}{\Gamma(1-\alpha)}) h(ia,t-s)ds\Big),$$

where $h(ia,.)$ is the unique continuous solution of the following fractional Riccati equation

$$D^\alpha h(ia,t) = \frac{1}{2}(-a^2 - ia) + (ia\rho\nu - \lambda)h(ia,t) + \frac{\nu^2}{2} h^2(ia,t), \quad I^{1-\alpha} h(ia,0) = 0, \tag{6}$$

We call Equation (6) a fractional Riccati equation. It can be also written in a Volterra form as follows,

$$h(ia,t) = \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)} \big(\frac{1}{2}(-a^2 - ia) + (ia\rho\nu - \lambda)h(ia,s) + \frac{\nu^2}{2}h^2(ia,s)\big)ds.$$

When $\alpha = 1$, we obtain a classical Riccati equation that can be solved explicitly leading to the celebrated Heston formula of the characteristic function, see [Hes93]. For $\alpha < 1$, (6) cannot be solved explicitly. However, we can apply the Adams scheme on the Volterra form of (6) to solve it numerically, see Section 5.1 in Chapter II.

Thanks to the semi-closed formula of the characteristic function, pricing vanilla options becomes an easy task in this model. Finally note that this formula has been extended in [AJLP17], to the general class of Volterra affine models in which the fractional kernel is replaced by an arbitrary one.

3 Part III: Hedging under the rough Heston model

Thanks to Result 5, we are now able to price derivatives under the rough Heston model. However in practice, the interest of pricing is limited if it does not go along with a hedging strategy.

Remarking that the conditional law of a rough Heston model is still a rough Heston dynamic, we are able to use again the Hawkes framework to compute the conditional characteristic function of the rough Heston model in Chapter III. This leads us to identify the state variables of the model, namely the underlying S_t and the forward variance curve $(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}$. As a result, any vanilla option can be perfectly hedged, in principle, using the underlying and the forward variance curve. In Chapter IV, we generalize this result for any Volterra Heston model and provide a general state space for the forward variance curve.

3.1 Conditional law of the rough Heston model

In order to derive a hedging strategy for a vanilla option with maturity $T > 0$ and payoff $f(S_T)$, we need to compute the dynamics of the process $(\mathbb{E}[f(S_T)|\mathscr{F}_t])_{t\in[0,T]}$, that is identifying the law of the rough Heston model (5) conditional on $\mathscr{F}_t$. We can show that the conditional law is still a rough Heston dynamic such that the mean reversion level θ becomes a time dependent one. Hence, it is convenient to extend the definition of the rough Heston model as follows:

$$dS_t = S_t\sqrt{V_t}dW_t$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta^0(s) - V_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\nu\sqrt{V_s}dB_s. \tag{7}$$

The parameters λ, V_0 and ν are positive, $\alpha \in (1/2, 1)$ and W and B are two Brownian motions with correlation ρ. The new mean reversion level θ^0 is allowed to be time dependent satisfying some regularity conditions (for technical reasons).

Assumption 6. *θ^0 is a deterministic function, continuous on $\mathbb{R}_+^*$ satisfying*

$$\forall u > 0; \quad \theta^0(u) \geq -\frac{V_0}{\lambda \Gamma(1-\alpha)} u^{-\alpha},$$

and

$$\forall \varepsilon > 0 \quad \exists K_\varepsilon > 0; \quad \forall u \in (0,1]; \quad \theta^0(u) \leq K_\varepsilon u^{-\frac{1}{2}-\varepsilon}.$$

Note that the fractional stochastic differential equation in (7) admits a unique weak solution, see Theorem 2 in Chapter III. We have the following result.

Result 6. *The law of the process $(S_t^{t_0}, V_t^{t_0})_{t\geq 0} = (S_{t+t_0}, V_{t+t_0})_{t \geq 0}$ is that of a rough Heston model with the following dynamic:*

$$dS_t^{t_0} = S_t^{t_0} \sqrt{V_t^{t_0}} dW_t^{t_0}, \quad S_0^{t_0} = S_{t_0}$$

$$V_t^{t_0} = V_{t_0} + \frac{1}{\Gamma(\alpha)} \int_0^t (t-u)^{\alpha-1} \lambda (\theta^{t_0}(u) - V_u^{t_0}) du + \frac{1}{\Gamma(\alpha)} \int_0^t (t-u)^{\alpha-1} \nu \sqrt{V_u^{t_0}} dB_u^{t_0},$$

with $(W_t^{t_0}, B_t^{t_0})_{t\geq 0} = (W_{t_0+t} - W_{t_0}, B_{t_0+t} - B_{t_0})_{t\geq 0}$ a two-dimensional Brownian motion with correlation ρ, independent of $\mathscr{F}_{t_0}$ and

$$\theta^{t_0}(u) = \theta^0(t_0+u) + \frac{\alpha}{\lambda\Gamma(1-\alpha)} \int_0^{t_0} (t_0 - v + u)^{-1-\alpha} (V_v - V_{t_0}) dv + \frac{(u+t_0)^{-\alpha}}{\lambda \Gamma(1-\alpha)} (V_0 - V_{t_0}),$$

which is an $\mathscr{F}_{t_0}$-measurable function satisfying Assumption 6.

3.2 Extending the characteristic function formula

In order to obtain the dynamic of the process $(\mathbb{E}[f(S_T)|\mathscr{F}_t])_{t\in[0,T]}$, we compute the conditional characteristic function of $X_T = \log(S_T/S_0)$. Thanks to Result 6, it is enough to extend the characteristic function formula stated in Result 5 by using again the Hawkes framework.

Result 7. *Let $t > 0$ and $z \in \mathbb{C}$. Assume*

$$\lambda - \rho\nu\Re(z) > 0, \quad a_-(t) < \Re(z) < a_+(t),$$

where

$$a_-(t) = \frac{\nu^2 - 2\rho\nu\psi(t) + \sqrt{\Delta(t)}}{2\nu^2(1-\rho^2)}, \quad a_+(t) = \frac{\nu^2 - 2\rho\nu\psi(t) - \sqrt{\Delta(t)}}{2\nu^2(1-\rho^2)},$$

with

$$\psi(t) = \lambda + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)}, \quad \Delta(t) = 4\nu^2\psi(t)^2 + \nu^4 - 4\rho\nu^3\psi(t).$$

Then we have

$$\mathbb{E}[(S_t)^{\Re(z)}] < \infty.$$

Furthermore,

$$R(z,t) = \mathbb{E}\left[\exp(z\log(S_t/S_0))\right]$$

is given by

$$\exp\Big(\int_0^t h(z,t-s)(\lambda\theta^0(s)+\frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\Big),$$

where $h(z,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha h(z,s)=\frac{1}{2}(z^2-z)+(z\rho\nu-\lambda)h(z,s)+\frac{\nu^2}{2}h(z,s)^2,\quad s\le t,\quad I^{1-\alpha}h(z,0)=0.$$

Note that the mean reversion level θ^0 is linked to the forward variance curve as follows:

$$\lambda\theta^0(t)=D^\alpha(\mathbb{E}[V_.]-V_0)(t)+\lambda\mathbb{E}[V_t].$$

So, we can write the characteristic function of the log-price as a functional of $(\mathbb{E}[V_s])_{s\ge0}$.

Result 8. *Let $t>0$ and $z\in\mathbb{C}$ satisfying the assumptions of Result 7. Then*

$$R(z,t)=\exp\Big(\int_0^t \chi(z,t-s)\mathbb{E}[V_s]ds\Big)$$

with

$$\chi(z,t)=\frac{1}{2}(z^2-z)+z\rho\nu h(z,t)+\frac{\nu^2}{2}h(z,t)^2,$$

with $h(z,.)$ the unique continuous solution of the fractional Riccati equation given in Result 7.

3.3 Hedging under the rough Heston model

We are now in position to compute the dynamic of a call price with maturity $T>0$ and strike $K>0$ under the rough Heston model

$$C_t=\mathbb{E}[(S_T-K)_+|\mathscr{F}_t].$$

Assuming that $\rho\le0$, Result 7 shows that there exists $a>1$ such that $\mathbb{E}[(S_t)^a]$ is finite for any time $t\ge0$. We are therefore able to use a Fourier inversion technique similar to [CM99] and obtain that

$$C_t=\frac{1}{2\pi}\int_{b\in\mathbb{R}}\hat{g}(-b)P_t^T(a+ib)db$$

with $\hat{g}(b)=\frac{e^{(1-a+ib)\log(K)}}{(ib-a)(ib-a+1)}$ and

$$P_t^T(a+ib)=\mathbb{E}[\exp\big((a+ib)\log(S_T)\big)|\mathscr{F}_t],$$

the conditional characteristic function process. Using the fact that conditional on $\mathscr{F}_t$, the law of (S,V) is still a rough Heston model and the characteristic function formula of Result 8, we deduce that

$$P_t^T(a+ib)=\exp\Big((a+ib)\log(S_t)+\int_0^{T-t}\chi(a+ib,T-t-s)\mathbb{E}[V_{s+t}|\mathscr{F}_t]ds\Big)$$

is a deterministic function of underlying S_t and the forward variance curve $(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\ge0}$. Hence there exists a deterministic functional C such that

$$C_t=C(T-t,S_t,(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\ge0}).$$

The following result gives an explicit hedging formula for C_t.

Result 9. *The functional C admits a derivative $\partial_S C$ according to the spot price S_t and a Fréchet derivative $\partial_V C$ according to the forward variance curve $(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}$. Furthermore, we have that*

$$C_t = C_0 + \int_0^t \partial_S C(T-u, S_u, \mathbb{E}[V_{\cdot+u}|\mathscr{F}_u]) dS_u + \int_0^t \partial_V C(T-u, S_u, \mathbb{E}[V_{\cdot+u}|\mathscr{F}_u]).(d\mathbb{E}[V_{\cdot+u}|\mathscr{F}_u]),$$

where $d\mathbb{E}[V_x|\mathscr{F}_u]$ is the Itô differential at time u of the martingale $M_u = \mathbb{E}[V_x|\mathscr{F}_u]$, $u \leq x$.

The importance of the last result is two-fold: It shows that we are able to hedge vanilla options in principle using the underlying and the forward variance curve and that the rough Heston model is Markovian in an infinite-dimensional space with state variables S_t and $(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}$. Of course, in practice, this strategy will be discretized and one will use liquid variance swaps. We can also hedge the forward variance curve risk by using one European option as the dynamics of the forward variance curve is produced by one Brownian noise.

3.4 The Markovian structure of the Volterra Heston model

After identifying the state variables of the rough Heston model, we investigate in Chapter IV a suitable general state space in which the forward variance curve lies. To do so, we extend the rough Heston model by replacing the fractional kernel $(t-s)^{\alpha-1}$ by an arbitrary one. This leads to the Volterra Heston model defined below.

$$dS_t = S_t\sqrt{V_t}dW_t$$

$$V_t = g_0(t) - \lambda\int_0^t K(t-s)V_s ds + \int_0^t K(t-s)\nu\sqrt{V_s}dB_s. \tag{8}$$

The parameters λ, V_0 and ν are positive and W and B are two Brownian motions with correlation ρ. The kernel $K \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}_+)$ is assumed to be completely monotone of the form

$$K(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t > 0,$$

where μ is a positive measure of locally bounded variation such that

$$\int_0^\infty (1\wedge(xh)^{-1/2})\mu(dx) \leq Ch^{(\gamma-1)/2}, \quad \int_0^\infty x^{-1/2}(1\wedge(xh))\mu(dx) \leq Ch^{\gamma/2}; \quad h > 0,$$

for some $\gamma \in (0,2]$ and positive constant C. The function g_0 is assumed to be continuous and satisfies

$$g_0(t) = \mathbb{E}[V_t] + \lambda\int_0^t K(t-s)\mathbb{E}[V_s]ds.$$

Hence, we can choose g_0 so that the model is consistent with the market forward variance curve.

3.4.1 Existence and uniqueness result

We start by looking for a general condition on g_0 to ensure the weak existence and uniqueness of a non-negative solution of the Volterra stochastic equation in (8) and hence the well-definition of the Volterra Heston model. Let $\mathcal{H}^{\gamma/2}$ be the set of locally Hölder continuous functions of any order strictly smaller than $\gamma/2$ and L be the resolvent of the first kind of K defined by the unique measure of locally bounded variation such that

$$L * K(t) = 1,$$

where $*$ is the convolution operator. We denote by Δ_h the following operator $\Delta_h : f \mapsto f(h+\cdot)$. We can show that $\Delta_h K * L$ is of bounded variation and we denote by $d(\Delta_h K * L)$ its associated measure. Following the same approach as in [AJLP17], we obtain the following result.

Result 10. *If g_0 belongs to the following admissible set:*

$$\mathcal{G}_K = \left\{g_0 \in \mathcal{H}^{\gamma/2}; g_0(0) \geq 0 \textit{ and } \quad \Delta_h g_0 - (\Delta_h K * L)(0) g_0 - d(\Delta_h K * L) * g_0 \geq 0, \textit{for any } h \geq 0\right\},$$

there exists a unique weak non-negative solution V of the Volterra stochastic equation (8).

Although the definition of $\mathcal{G}_K$ seems abstract, we can show that it contains many usual parameterizations of the forward variance curve, see Example IV.1 in Chapter IV.

3.4.2 Markovian structure

Following the same approach as in Chapter III, we identify the conditional law of the Volterra Heston model. More precisely, we obtain that the law of $(S_t^{t_0}, V_t^{t_0})_{t\geq 0} = (S_{t+t_0}, V_{t+t_0})_{t\geq 0}$ is that of a Volterra Heston model with the following dynamics:

$$dS_t^{t_0} = S_t^{t_0}\sqrt{V_t^{t_0}}dW_t^{t_0}, \quad S_0^{t_0} = S_{t_0}$$

$$V_t^{t_0} = g_{t_0}(t) - \lambda\int_0^t K(t-u)V_u^{t_0}du + \int_0^t K(t-u)\nu\sqrt{V_u^{t_0}}dB_u^{t_0},$$

with $(W_t^{t_0}, B_t^{t_0})_{t\geq 0} = (W_{t_0+t} - W_{t_0}, B_{t_0+t} - B_{t_0})_{t\geq 0}$ a two-dimensional Brownian motion with correlation ρ, independent of $\mathcal{F}_{t_0}$ and g_{t_0} linked to the forward variance curve observed at time t_0,

$$g_{t_0}(t) = \mathbb{E}[V_t^{t_0}|\mathcal{F}_{t_0}] + \lambda\int_0^t K(t-s)\mathbb{E}[V_s^{t_0}|\mathcal{F}_{t_0}]ds.$$

Hence, similarly to Result 6, the Volterra Heston model is Markovian with state variables the underlying S_t and g_t which is linked to the forward variance curve $(\mathbb{E}[V_{s+t}|\mathcal{F}_t])_{s\geq 0}$. Moreover, we are able to provide the state space in which the process $(g_t)_{t\geq 0}$ evolves.

Result 11. *$\mathcal{G}_K$ is stochastically invariant according to $(g_t)_{t\geq 0}$, that is*

$$g_t \in \mathcal{G}_K, \quad t \geq 0.$$

4 Part IV: The rough Heston model in practice

We summarize in Chapter V the results obtained under the rough Heston model and show the excellent fit of this model to the SPX volatility surface. We also study in Chapter VI the consistency of this model with a feature raised in [Zum09] concerning the time reversal asymmetry of financial time series.

4.1 Fitting the volatility surface

In Parts II and III, we build a rough version of the Heston model that appears naturally as a limit of a microscopic price and show that it enjoys tractable formulas for derivatives pricing and hedging. In Chapter V, we apply a calibration procedure on the SPX volatility surface and show the amazing fit of the model to the actual volatility surface. We also suggest, by a moments-matching procedure, an approximation method to compute instantaneously the implied volatility for a given expiry T. More precisely, we approximate the characteristic function by the one of a classical Heston model with flat forward variance curve given by $\frac{1}{T}\int_0^T \mathbb{E}[V_s]ds$, correlation ρ and volatility of volatility parameter given by

$$\widetilde{\nu}(T) = \sqrt{\frac{3}{2H+2}}\,\frac{\nu}{\Gamma\left(H+\frac{3}{2}\right)}\,\frac{1}{T^{\frac{1}{2}-H}}.$$

4.2 Consistency with the time reversal asymmetry

In [CB14, Zum09], it is observed empirically that the time reversal symmetry is violated in financial time series. This is the so-called Zumbach's effect. It is done by studying the time series of daily asset log-returns $(r_t)_{t\geq 0}$ and daily realized volatility $(v_t)_{t\geq 0}$. More precisely, we observe that the empirical correlation between $r_{t-\delta}^2$ and v_t is greater that the empirical correlation between r_t^2 and $v_{t-\delta}$ for any $\delta > \delta_0 = 1$ day. Under the rough Heston model (7), the daily return is given by[1]

$$r_t = \int_{t-\delta_0}^{t} \sqrt{V_s}dW_s,$$

and the daily realized volatility is

$$v_t = \int_{t-\delta_0}^{t} V_s ds.$$

In Chapter VI, we show that the rough Heston model exhibits this phenomenon by computing

$$C(k,t) = Cov(v_{t+k\delta_0}, r_t^2) - Cov(r_{t+k\delta_0}^2, v_t), \quad t > \delta_0, \quad k \in \mathbb{N}.$$

Result 12. *For small δ_0, $C(k,t)$ is given approximatively by*

$$C(k,t) \underset{\delta_0\to 0}{\sim} 2(\rho\nu)^2\delta_0^{2\alpha+1}g(k)\mathbb{E}[V_t], \quad k\in\mathbb{N}, \quad t>0,$$

with $g(k) = \frac{1}{\Gamma(\alpha+1)^2}\int_0^1 \left((k+s)^\alpha - (k+s-1)^\alpha\right)(1-s)^\alpha ds$ *and* $\alpha = H+1/2$. *In particular* $C(k,t) > 0$.

[1]For simplicity, we take only the martingale part of the log-return.

We also compute the Zumbach's effect in the stationary regime $Z(k)$ defined by

$$Z(k) = \lim_{t\to\infty} Corr(v_{t+k\delta_0}, r_t^2) - Corr(r_{t+k\delta_0}^2, v_t).$$

Result 13. *Consider the case where $\mathbb{E}[V_t]$ converges to a limiting variance $\bar{V}_\infty$ when t goes to infinity. Then, for small δ_0,*

$$Z(k) \sim \frac{2(\rho\nu)^2\sqrt{\bar{V}_\infty}}{\sqrt{\frac{\nu^2}{\lambda^2}\int_0^\infty f^{\alpha,\lambda}(s)^2 ds}\sqrt{2\bar{V}_\infty + \frac{6\nu^2}{\lambda^2}\int_0^\infty f^{\alpha,\lambda}(s)^2 ds}}\,\delta_0^{2\alpha-1} g(k),$$

where $f^{\alpha,\lambda}$ is the Mittag-Leffler density function, $\alpha = H+1/2$ and $g(k) = \frac{1}{\Gamma(\alpha+1)^2}\int_0^1 ((k+s)^\alpha - (k+s-1)^\alpha)(1-s)^\alpha ds$.

Note that the Zumbach's effect is of order $\delta_0^{2\alpha-1} = \delta_0^{2H}$. Hence as δ_0 is small, this effect is negligible when $H = 1/2$, and this effect becomes more important when the volatility is rough, i.e when H is close to zero.

5 Part V: Short-term behavior of the at-the-money implied volatility under rough volatility

While the Hawkes framework enabled us to to understand the pricing and hedging under the rough Heston model, this procedure cannot be applied for an arbitrary rough volatility model. We show in Chapter VII that asymptotic formulas for short maturity for the at-the-money skew and curvature can be obtained under a general class of stochastic volatility (that includes usual rough volatility models). This is done by applying a small-time Edgeworth expansion of the density of the asset price.

5.1 Model

We consider a general model such that the log-price is adapted to a filtration $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$ and is given by

$$dX_t = -\frac{1}{2}V_t dt + \sqrt{V_t} dW_t,$$

where the variance process V is positive and adapted to a filtration $\mathbb{G} = (\mathscr{G}_t)_{t\geq 0}$ smaller than $\mathbb{F}$. We assume that the Brownian motion is decomposed as $W_t = \rho_t B_t + \sqrt{1-\rho_t^2} B'_t$, with B' independent of $\mathbb{G}$ and B is a $\mathbb{G}$-Brownian motion. Denoting by $p(K,\theta)$ the put option price with maturity θ and strike K, we have

$$\frac{p(Fe^{z\sigma_0(\theta)},\theta)}{F\sigma_0(\theta)} = \int_{-\infty}^z \mathbb{P}(Z_\theta \leq \zeta) e^{\zeta\sigma_0(\theta)} d\zeta,$$

where $F = e^{X_0}$ is the forward price, $\sigma_0(\theta) = \sqrt{\int_0^\theta \mathbb{E}[V_s] ds}$ and

$$Z_\theta = -\frac{1}{2\sigma_0(\theta)}\langle M\rangle_\theta + \frac{1}{\sigma_0(\theta)} M_\theta, \quad M_\theta = \int_0^\theta \sqrt{V_t} \mathrm{d}B_t, \quad \langle M\rangle_\theta = \int_0^\theta V_t \mathrm{d}t.$$

In order to apply a small-time expansion of the option prices, we need to make the following technical assumption that settles the asymptotic behavior of the model.

Assumption 7. *There exists a family of random vectors*

$$\left\{(M_\theta^{(0)}, M_\theta^{(1)}, M_\theta^{(2)}, M_\theta^{(3)}); \theta \in (0,1)\right\}$$

such that

1. *the law of $M_\theta^{(0)}$ is standard normal for all $\theta > 0$,*

2. $$\sup_{\theta \in (0,1)} \|M_\theta^{(i)}\|_p < \infty, \quad i = 1,2,3$$

   for all $p > 0$ and

3. *for some $H \in (0, 1/2]$ and $\epsilon \in (0, H)$,*

$$\lim_{\theta \to 0} \theta^{-2H-2\epsilon} \left\| \frac{M_\theta}{\sigma_0(\theta)} - M_\theta^{(0)} - \theta^H M_\theta^{(1)} - \theta^{2H} M_\theta^{(2)} \right\|_{1+\epsilon} = 0,$$
$$\lim_{\theta \to 0} \theta^{-H-2\epsilon} \left\| \frac{\langle M \rangle_\theta}{\sigma_0(\theta)^2} - 1 - \theta^H M_\theta^{(3)} \right\|_{1+\epsilon} = 0.$$

Furthermore, we assume the existence of the derivatives

$$a_\theta^{(i)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \left\{ E_0[M_\theta^{(i)} | M_\theta^{(0)} = x] \phi(x) \right\}, \quad i = 1,2,3,$$
$$b_\theta(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \left\{ E_0[M_\theta^{(1)} | M_\theta^{(0)} = x] \phi(x) \right\}$$
$$c_\theta(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \left\{ E_0[|M_\theta^{(1)}|^2 | M_\theta^{(0)} = x] \phi(x) \right\}$$

in the Schwartz space (i.e., the space of the rapidly decreasing smooth functions), where ϕ is the standard normal density. Finally we assume that

$$\sup_{\theta \in (0,1)} \left\| \frac{1}{\theta} \int_0^\theta V_t \mathrm{d}t \right\|_p < \infty, \quad \sup_{\theta \in (0,1)} \left\| \left\{ \frac{1}{\theta} \int_0^\theta V_t (1 - \rho_t^2) \mathrm{d}t \right\}^{-1} \right\|_p < \infty,$$

where $\|\cdot\|_p$ denotes the $\mathbb{L}^p$-norm under the probability measure $\mathbb{P}$.

Note that regular stochastic volatility models satisfy Assumption 7, see Section 2.2 of Chapter VII.

5.2 Asymptotics results

Applying Assumption 7, we obtain an asymptotic expansion of the density of Z_θ,.

Result 14. *Under Assumption 7, the density p_θ of Z_θ satisfies*

$$\sup_{x\in\mathbb{R}}(1+x^2)^\alpha|p_\theta(x)-q_\theta(x)|=o(\theta^{2H})$$

as $\theta\to 0$ for any $\alpha\in\mathbb{N}$, where

$$\begin{aligned}q_\theta(x)=&\phi(x)-\theta^H a_\theta^{(1)}(x)-\theta^{2H}a_\theta^{(2)}(x)-\frac{\sigma_0(\theta)}{2}(x\phi(x)-\theta^H a_\theta^{(3)}(x))\\&+\frac{\theta^{2H}}{2}c_\theta(x)-\frac{\theta^H\sigma_0(\theta)}{2}b_\theta(x)+\frac{\sigma_0(\theta)^2}{8}(x^2-1)\phi(x).\end{aligned}$$

Under usual stochastic volatility models, we can in general approximate q_θ by $\widetilde{q}_\theta$ such that

$$\sup_{x\in\mathbb{R}}(1+x^2)^\alpha|q_\theta(x)-\widetilde{q}_\theta(x)|=o(\theta^{2H}),$$

with

$$\begin{aligned}\widetilde{q}_\theta(x)=&\phi\left(x+\frac{\sigma_0(\theta)}{2}\right)\left\{1+\kappa_3(\theta)\left(H_3\left(x+\frac{\sigma_0(\theta)}{2}\right)-\sigma_0(\theta)H_2\left(x+\frac{\sigma_0(\theta)}{2}\right)\right)\theta^H\right\}\\&+\phi(x)\left(\kappa_4(\theta)H_4(x)+\frac{\kappa_3(\theta)^2}{2}H_6(x)\right)\theta^{2H},\end{aligned}$$

where κ_3 and κ_4 are bounded functions and H_k denotes the kth Hermite polynomial. This enables us to compute the asymptotic behavior of the put price $p(Fe^{z\sigma_0(\theta)},\theta)$ for small θ, leading to the following expansion of the implied volatility.

Result 15. *For any $z\in\mathbb{R}$,*

$$\begin{aligned}&\sigma_{BS}(\sqrt{\theta}z,\theta)\\&=\kappa_2(\theta)\left\{1+\kappa_3(\theta)\left(\frac{z}{\kappa_2(\theta)}+\frac{\kappa_2(\theta)\sqrt{\theta}}{2}\right)\theta^H+\left(\frac{3\kappa_3^2(\theta)}{2}-\kappa_4(\theta)+(\kappa_4(\theta)-3\kappa_3^2(\theta))\frac{z^2}{\kappa_2^2(\theta)}\right)\theta^{2H}\right\}+o(\theta^{2H}),\end{aligned}$$

where $\kappa_2(\theta)=\sigma_0(\theta)/\sqrt{\theta}$ and $\sigma_{BS}(k,\theta)$ denotes the implied volatility with log-moneyness k and maturity θ. This leads to the following asymptotics of the at-the-money skew and curvature

$$\begin{aligned}\partial_k\sigma_{BS}(0,\theta)&=\kappa_3(\theta)\theta^{H-1/2}+o(\theta^{2H-1/2}),\\\partial_k^2\sigma_{BS}(0,\theta)&=2\frac{\kappa_4(\theta)-3\kappa_3(\theta)^2}{\kappa_2(\theta)}\theta^{2H-1}+o(\theta^{2H-1}).\end{aligned}$$

5.3 Application to the rough Bergomi model

We apply these results to the rough Bergomi model introduced in [BFG16], where the variance process is defined as follows

$$V_t = \mathbb{E}[V_t] \exp\left\{ \eta_H \sqrt{2H} \int_0^t (t-s)^{H-1/2} \mathrm{d}B_s - \frac{\eta_H^2}{2} t^{2H} \right\},$$

where $d\langle W, B\rangle_t = \rho dt$. We can check then that Assumption 7 is satisfied and that $\widetilde{q}_\theta$ can be computed explicitly with

$$\kappa_3(\theta) = \rho \eta_H \sqrt{\frac{H}{2}} \frac{1}{\theta^H \sigma_0(\theta)^3} \int_0^\theta \exp\left\{ -\frac{\eta_H^2}{8} t^{2H} \right\} \int_0^t (t-s)^{H-1/2} \sqrt{\mathbb{E}[V_s]} \mathrm{d}s \mathbb{E}[V_t] \mathrm{d}t,$$
$$\kappa_4(\theta) = \frac{(1+2\rho^2)\eta_H^2 H}{(2H+1)^2(2H+2)} + \frac{\rho^2 \eta_H^2 H \beta(H+3/2, H+3/2)}{2(H+1/2)^2}.$$

where β is the beta function.

6 Part VI: Markovian approximation of rough volatility models

In Chapter VIII, we consider the following general class of rough volatility models:

$$dS_t = S_t \sqrt{V_t} dW_t, \quad t \in [0, T],$$

$$V_t = V_0 + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} \lambda(\theta^0(s) - V_s) ds + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} \sigma(V_s) dB_s. \quad (9)$$

The parameters λ, V_0 are positive, $H \in (0, 1/2)$ is the Hurst parameter[2], σ is a deterministic function, θ^0 is a mean reversion level allowed to be time dependent and W and B are two Brownian motions with correlation ρ. Note that the rough Heston model (7) is a particular case of (9) with $\sigma(x) = \nu\sqrt{x}$. The weak existence of such model is guaranteed by assuming that θ^0 satisfies Assumption 6 and that σ is a continuous function with linear growth such that $\sigma(0) = 0$. In that case, the variance process V admits Hölder continuous paths of any order strictly less than H.

Due to the fractional kernel $(t-s)^{H-1/2}$, the variance process is neither Markovian nor a semi-martingale. In the spirit of Parts II and III, we would like to overcome these difficulties in order to manage the pricing and hedging of derivatives by looking for a tractable approximation of (9). We expect from this approximation to display a Markovian structure such that the variance process is a semi-martingale.

[2] Note the change of notations compared to (5) in the fractional kernel power where α is replaced by $H + 1/2$. This is to be consistent with the notations of Chapter VIII.

6.1 Multi-factor approximation of rough volatility models

Inspired by [CC98, CCM00, HS15, Mur11], the key-idea of our multi-factor approximation comes from the observation that the fractional kernel $K(t) = \frac{t^{H-1/2}}{\Gamma(H+1/2)}$ is the Laplace transform of a non-negative measure μ:

$$K(t) = \int_0^\infty e^{-\gamma t}\mu(d\gamma); \quad \mu(d\gamma) = \frac{\gamma^{-H-\frac{1}{2}}}{\Gamma(H+1/2)\Gamma(1/2-H)}d\gamma.$$

Applying a stochastic Fubini Theorem, we obtain that

$$V_t = g(t) + \int_0^\infty V_t^\gamma \mu(d\gamma), \quad t \in [0,T],$$

with

$$dV_t^\gamma = (-\gamma V_t^\gamma - \lambda V_t)dt + \sigma(V_t)dB_t, \quad V_0^\gamma = 0, \quad \gamma \geq 0,$$

and

$$g(t) = V_0 + \int_0^t K(t-s)\theta^0(s)ds.$$

Heuristically, this exhibits an infinite-dimensional Markovian structure of the variance process where the state variables are given by the factors $(V_t^\gamma)_{\gamma\geq 0}$. In order to reduce the number of state variables, we shall approximate the measure μ by a weighted sum of Dirac measures

$$\mu^n = \sum_{i=1}^n c_i^n \delta_{\gamma_i^n}, \quad n \geq 1,$$

leading to the following approximation $V^n = (V_t^n)_{t\leq T}$ of the variance process V:

$$V_t^n = g^n(t) + \sum_{i=1}^n c_i^n V_t^{n,i}, \quad t \in [0,T], \tag{10}$$

$$dV_t^{n,i} = (-\gamma_i^n V_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n)dB_t, \quad V_0^{n,i} = 0,$$

where

$$g^n(t) = V_0 + \int_0^t K^n(t-u)\theta^0(u)du, \tag{11}$$

and

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

Before discussing the accuracy of this approximation, we first prove the existence and uniqueness of (10) by writing it in a Volterra stochastic equation form:

$$V_t^n = g^n(t) + \int_0^t K^n(t-s)\left(-\lambda V_s^n ds + \sigma(V_s^n)dB_s\right), \quad t \in [0,T]. \tag{12}$$

Consequently, we adapt the arguments in [AJLP17] to prove the weak existence of a non-negative solution of (12) and those in [YW71] to obtain the strong uniqueness under the following assumption.

Assumption 8. *We assume that θ^0 is a non-negative function satisfying Assumption 6 and that σ is η-Hölder continuous with $\sigma(0) = 0$ and $\eta \in [1/2, 1]$.*

This leads us to the following definition of the multi-factor approximation

$$dS_t^n = S_t^n \sqrt{V_t^n} dW_t, \quad t \in [0, T],$$

with V^n given by (10). This model is Markovian with $n+1$ state variables which are the spot price S^n and the factors of the variance process $V^{n,i}$ for $i \in \{1, \ldots, n\}$.

6.2 Convergence of the multi-factor approximation

We now discuss the choice of the positive weights $(c_i^n)_{1 \le i \le n}$ and mean reversion coefficients $\gamma_1^n < \cdots < \gamma_n^n$, which is crucial for the accuracy of the approximation. From (12), we expect that V^n converges to V if the kernel K^n is close to K for large n. Hence inspired by [CCM00], we look for a condition on the weights and mean-reversions such that

$$\int_0^T \left| K^n(s) - K(s) \right|^2 ds \to 0,$$

holds as n goes to infinity. This is done by taking

$$c_i^n = \int_{\eta_{i-1}^n}^{\eta_i^n} \mu(d\gamma), \quad \gamma_i^n = \frac{1}{c_i^n} \int_{\eta_{i-1}^n}^{\eta_i^n} \gamma \mu(d\gamma), \quad i \in \{1, \ldots, n\}.$$

where $(\eta_i^n)_{0 \le i \le n}$ are auxiliary parameters with $\eta_0^n = 0$ and $\eta_{i-1}^n \le \gamma_i^n \le \eta_i^n$ for $i \in \{1, \ldots, n\}$. In that case, we obtain the convergence of the approximated kernel K^n to K under the following assumption.

Assumption 9. *We assume that the auxiliary parameters $(\eta_i^n)_{0 \le i \le n}$ satisfy*

$$\eta_n^n \to \infty, \quad \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma_i^n - \gamma)^2 \mu(d\gamma) \to 0,$$

as n goes to infinity.

The next result shows that under this choice of the weights and mean-reversions, the multi-factor approximation is accurate and converges to the proper rough volatility model.

Result 16. *Under Assumptions 8 and 9, The sequence of multi-factor models $(S^n, V^n)_{n \ge 1}$ is tight for the uniform topology and any point limit (S, V) is the rough volatility model* (9).

This result is obtained by studying the stability of general d-dimensional stochastic Volterra equations, see Section 3.4 in Chapter VIII.

In the same spirit as Result 16, Markovian lifts of general affine rough volatility models have been developed in [CT18].

6.3 Application to the rough Heston model

We now come back to the rough Heston case, *i.e* $\sigma(x) = \nu\sqrt{x}$. Results 5 and 7 provide the formula of the log-price characteristic function. It can be rewritten as follows:

$$\exp\left(\int_0^T F(z, h(z, T-s))g(s)ds\right),$$

where $z \in \mathbb{C}$ with $\Re(z) \in [0,1]$, $F(z,x) = \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2$ and $h(z,\cdot)$ is the unique solution of the fractional Riccati equation given in Result 8. This fractional Riccati equation can be solved numerically through the Adam schemes presented in Section 5.1 in Chapter II. We show that the multi-factor approximation gives rise to another natural numerical scheme.

6.3.1 Multi-factor scheme for the fractional Riccati equations

Considering the multi-factor approximation $(S^n, V^n)_{n\geq 1}$ of the rough Heston model, we can use the characteristic function formula shown in [AJLP17] that generalizes Result 5 for Volterra affine models to get

$$\mathbb{E}[\exp(z\log(S_T^n/S_0))] = \exp(\int_0^T F(z, h^n(z, T-s))g^n(s)ds),$$

where $h^n(z,.)$ is the unique solution of the Volterra Riccati equation,

$$h^n(z,t) = \int_0^t K^n(t-s)F(z, h^n(z,s))ds.$$

As K^n is close to K under Assumption 9 for large n, we expect the convergence of $h^n(z,.)$ to the solution of the fractional Riccati equation $h(z,.)$.

Result 17. *There exists a positive constant C such that, for any $a \in [0,1]$, $b \in \mathbb{R}$ and $n \geq 1$,*

$$\sup_{t\in[0,T]} |h^n(a+ib,t) - h(a+ib,t)| \leq C(1+b^4)\int_0^T |K^n(s) - K(s)|ds.$$

This result suggests a new numerical method for the computation of the fractional Riccati solution $h(z,\cdot)$. In fact, defining

$$h^{n,i}(z,t) = \int_0^t e^{-\gamma_i^n(t-s)}F(z, h^n(z,s))ds, \quad i \in \{1,\dots,n\},$$

we get

$$h^n(z,t) = \sum_{i=1}^n c_i^n h^{n,i}(z,t),$$

and $(h^{n,i}(z,\cdot))_{1\leq i\leq n}$ solves the following n-dimensional system of ordinary Riccati equations

$$\partial_t h^{n,i}(z,t) = -\gamma_i^n h^{n,i}(z,t) + F(z, h^n(z,t)), \quad h^{n,i}(z,0) = 0, \quad i \in \{1,\dots,n\}.$$

This n-dimensional Riccati equation can be solved numerically by usual finite difference methods leading to $h^n(z,\cdot)$ as an approximation of the fractional Riccati solution. Numerical illustrations are given in Section 4.2 in Chapter VIII, where we compare the convergence speed of this numerical for different choices of the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversion parameters $(\gamma_i^n)_{1\leq i\leq n}$.

6.3.2 Upper bound for call prices error

Using Result 17 and a Fourier transform method, we compute an upper bound of the difference between the call price $C^n(T,k)$ under the multi-factor approximation and the call price $C(T,k)$ under the rough Heston model, where k denotes the log-strike and T the maturity. In order to obtain an explicit upper bound, we consider a multi-factor approximation (S^n, V^n) where g^n initially defined by (11) is replaced by

$$g^n(t) = \int_0^t K^n(t-s)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2-H)} + \theta^0(s)\big)ds.$$

Note that in that case, the convergence of $(S^n, V^n)_{n\geq 1}$ stated in Result 16 does not hold anymore. However, we may show the convergence in law for the Skorokhod topology of $(S^n, \int_0^\cdot V_s^n ds)_{n\geq 1}$ to $(S, \int_0^\cdot V_s ds)$.

Result 18. *There exists a positive constant $c > 0$ such that*

$$|C(k,T) - C^n(k,T)| \leq c\int_0^T |K(s) - K^n(s)|ds, \quad n \geq 1.$$

7 Part VII: Optimal make-take fees for market making regulation

In Chapter IX, we consider the problem of an exchange aiming at attracting liquidity on its platform. This is done by creating an incentive for the market maker to reduce the spread through a compensation ξ proposed by the exchange. The aim of Chapter IX is to find the optimal contract ξ that maximizes the utility of the exchange. We solve this problem using a principal-agent framework developed in [Höl79, Mir74]. We argue that this optimal policy leads to higher quality liquidity and lower trading costs for investors.

7.1 The model

Let $T > 0$ be a final time horizon. We consider a market where there is only one market maker who has a view on the efficient price of the asset S_t given by a Bachelier dynamic:

$$S_t = S_0 + \sigma W_t, \quad t \in [0,T], \tag{13}$$

with $S_0 > 0$, W a Brownian motion and $\sigma > 0$ the volatility of the price. This market maker sets the bid and ask prices

$$P_t^b = S_t - \delta_t^b \quad \text{and} \quad P_t^a = S_t + \delta_t^a.$$

The arrival of the ask (resp. bid) market orders is modeled by a point process $(N_t^a)_{t\in[0,T]}$ (resp. $(N_t^b)_{t\in[0,T]}$) with intensity $(\lambda_t^a)_{t\in[0,T]}$ (resp. $(\lambda_t^b)_{t\in[0,T]}$). Hence, assuming that the volume of market orders is constant and equal to unity, the inventory of the market maker is

$$Q_t = N_t^b - N_t^a, \quad t \in [0,T].$$

We impose a critical absolute inventory $\bar{q} \in \mathbb{N}$ above which the market maker stops quoting on the ask or bid side. Moreover, we take the intensity of buy (resp. sell) market orders arrival as a decreasing function of the ratio between the extra cost of each trade paid by the market taker compared to the efficient price. This extra cost is the sum of the spread δ_t^a (resp. δ_t^b) imposed by the market maker and the transaction cost $c > 0$ collected by the exchange for each order. This leads us to assume that

$$\lambda_t^a = \lambda(\delta_t^a)\mathbb{1}_{\{Q_t > -\bar{q}\}}, \quad \text{and} \quad \lambda_t^b = \lambda(\delta_t^b)\mathbb{1}_{\{Q_t < \bar{q}\}}, \quad \text{with} \quad \lambda(x) = Ae^{-k\frac{(x+c)}{\sigma}},$$

for fixed positive constants A and k.

7.1.1 Admissible controls and market maker's problem

The market maker controls the spread $(\delta_t)_{t\in[0,T]} = (\delta_t^a, \delta_t^b)_{t\in[0,T]}$ which can be any predictable process that is uniformly bounded by a positive constant δ_∞ that will be fixed later to a sufficiently large value. We denote by $\mathscr{A}$ the set of such admissible controls and by $\mathbb{P}^\delta$ the probability measure associated to the control δ under which S follows (13) and

$$\widetilde{N}_t^{\delta,a} = N_t^a - \int_0^t \lambda(\delta_r^a)\mathbb{1}_{\{Q_r > -\bar{q}\}} dr, \quad \widetilde{N}_t^{\delta,b} = N_t^b - \int_0^t \lambda(\delta_r^b)\mathbb{1}_{\{Q_r < \bar{q}\}} dr,$$

are martingales. Thanks to the uniform boundedness of the controls, we show that all the probability measures $\mathbb{P}^\delta$ indexed by $\delta \in \mathscr{A}$ are equivalent.

Under the control $\delta \in \mathscr{A}$, the market maker profit and loss PL_t^δ at time t is the sum of the cash flow earned from each market order

$$X_t^\delta = \int_0^t P_r^a dN_r^a - \int_0^t P_r^b dN_r^b,$$

and his inventory risk $Q_t S_t$. Furthermore, we consider that the exchange is remunerated for each market order arrival and so aims at keeping the market liquid. Thus, we assume that it proposes to the market maker a contract, defined by an $\mathscr{F}_T$-measurable random variable ξ, in order to create an incentive to attract liquidity on the platform by reducing his spread. Hence, in addition to the realized profit and loss, the market maker receives this compensation ξ at the final time T, and chooses the optimal spread by solving the following problem

$$V_{\mathrm{MM}}(\xi) = \sup_{\delta\in\mathscr{A}} J_{\mathrm{MM}}(\delta,\xi) \quad \text{where} \quad J_{\mathrm{MM}}(\delta,\xi) = \mathbb{E}^\delta\Big[-e^{-\gamma(\xi + \mathrm{PL}_T^\delta - \mathrm{PL}_0^\delta)}\Big], \tag{14}$$

where $\gamma > 0$ is the absolute risk aversion parameter of the CARA utility function of the market maker. For each compensation ξ, we show that there exists a unique optimal response $\hat{\delta}(\xi) = (\hat{\delta}^a(\xi), \hat{\delta}^b(\xi))$ of the market marker.

7.1.2 Admissible contracts and exchange's problem

The exchange is remunerated by $c > 0$ for each market order and pays the contract ξ at the final time T. Hence, its profit and loss during the time interval $[0, T]$ is

$$c(N_T^a - N_0^a + N_T^b - N_0^b) - \xi.$$

Thus the exchange optimally chooses the contract to maximize its CARA utility function with absolute risk aversion parameter $\eta > 0$,

$$V_0^E = \sup_{\xi \in \mathscr{C}} \mathbb{E}^{\hat{\delta}(\xi)}\Big[- e^{-\eta(c(N_T^a - N_0^a + N_T^b - N_0^b) - \xi)}\Big], \tag{15}$$

where $\mathscr{C}$ is the set of admissible contracts. It is chosen to be the set of any $\mathscr{F}_T$-measurable random variable satisfying some integrability conditions to ensure the well-definition of the problems (14) and (15), and $V_{\mathrm{MM}}(\xi) > R$, since we consider a market maker that accepts only contracts with optimal utility above a threshold utility R.

7.2 Solving the market maker's problem

We start by solving the problem (14) of the market maker facing an arbitrary contract $\xi \in \mathscr{C}$ proposed by the exchange. In order to do so, we first solve (14) when ξ has the following representation

$$\xi = Y_T^{Y_0,Z} = Y_0 + \int_0^t Z_r^a dN_r^a + Z_r^b dN_r^b + Z_r^S dS_r + \Big(\frac{1}{2}\gamma\sigma^2 (Z_r^S + Q_r)^2 - H(Z_r, Q_r)\Big) dr, \tag{16}$$

where $Y_0 \in \mathbb{R}$ and $Z = (Z^a, Z^b, Z^S)$ belongs to the set $\mathcal{Z}$ of predictable processes satisfying some integrability conditions. Here, $H(z,q)$ denotes the Hamiltonian of (14) defined by

$$H(z,q) = \sup_{|\delta^a| \vee |\delta^b| \le \delta_\infty} h(\delta, z, q),$$

with

$$h(\delta, z, q) = \frac{1 - e^{-\gamma(z^a + \delta^a)}}{\gamma} \lambda(\delta^a) \mathbb{1}_{\{q > -\bar{q}\}} + \frac{1 - e^{-\gamma(z^b + \delta^b)}}{\gamma} \lambda(\delta^b) \mathbb{1}_{\{q < \bar{q}\}},$$

for any $\delta = (\delta^a, \delta^b) \in [-\delta_\infty, \delta_\infty]^2$ and $z = (z^S, z^a, z^b) \in \mathbb{R}^3$ and $q \in \mathbb{Z}$.

When $\xi = Y_T^{Y_0,Z}$, we can show using Itô's formula on

$$U_T^\delta = -\exp\Big(-\gamma(Y_T^{Y_0,Z} + \mathrm{PL}_T^\delta - \mathrm{PL}_0^\delta)\Big),$$

that $\mathbb{E}^\delta[U_T^\delta]$ is optimal when $\delta = \hat{\delta}(\xi) = (\hat{\delta}^a(\xi), \hat{\delta}^b(\xi))$ given by

$$\hat{\delta}_t^a(\xi) = \Delta(Z_t^a),\ \hat{\delta}_t^b(\xi) = \Delta(Z_t^b),\ \text{where } \Delta(z) = (-\delta_\infty) \vee \Big\{ - z + \frac{1}{\gamma}\log\Big(1 + \frac{\sigma\gamma}{k}\Big)\Big\} \wedge \delta_\infty. \tag{17}$$

In that case, the optimal utility is $V_{MM}(\xi) = -e^{-\gamma Y_0}$.

The following result states that any admissible contract $\xi \in \mathscr{C}$ admits a unique representation given by (16).

Result 19. *For any admissible contract ξ, there exist a unique $Y_0 \in \mathbb{R}$ and $Z \in \mathcal{Z}$ such that $\xi = Y_T^{Y_0,Z}$. Under this representation, $V_{MM}(\xi) = -e^{-\gamma Y_0}$ and the optimal spread $\hat{\delta}(\xi)$ is given by* (17). *In particular, the set of admissible contracts $\mathscr{C}$ coincides with*

$$\Xi = \Big\{ Y_T^{Y_0,Z} :\ Z \in \mathcal{Z},\ \textit{and}\ \ Y_0 \ge \hat{Y}_0 = \frac{-1}{\gamma}\log(-R)\Big\}.$$

7.3 Designing the optimal contract

Using Result 19, the exchange's problem (15) reduces to

$$V_0^E = \sup_{Y_0 \geq \hat{Y}_0} \sup_{Z \in \mathcal{Z}} \mathbb{E}^{\hat{\delta}(Y_T^{Y_0,Z})} \Big[-e^{-\eta \big(c(N_T^a - N_0^a + N_T^b - N_0^b) - Y_T^{Y_0,Z} \big)} \Big].$$

The objective function above being decreasing in Y_0, we get that

$$V_0^E = e^{\eta \hat{Y}_0} \sup_{Z \in \mathcal{Z}} \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})} \Big[-e^{-\eta \big(c(N_T^a - N_0^a + N_T^b - N_0^b) - Y_T^{0,Z} \big)} \Big].$$

Hence, the exchange's problem (15) is reduced to a classical stochastic control problem. The HJB equation of this control problem is given by

$$\begin{cases} \partial_t v(t,q) + \frac{\gamma \eta^2 \sigma^2}{2(\gamma+\eta)} q^2 v(t,q) - C v(t,q) \Big[\mathbb{1}_{\{q > -\bar{q}\}} \big(\frac{v(t,q)}{v(t,q-1)} \big)^{\frac{k}{\sigma\eta}} + \mathbb{1}_{\{q < \bar{q}\}} \big(\frac{v(t,q)}{v(t,q+1)} \big)^{\frac{k}{\sigma\eta}} \Big] = 0, \\ v(T,q) = -1, \end{cases} \tag{18}$$

with $C = A \frac{\sigma\eta}{k} \exp\Big(-\frac{k}{\sigma\gamma} \log(1 + \frac{\sigma\gamma}{k}) + (1 + \frac{k}{\sigma\eta}) \log\big(1 - \frac{\sigma^2 \gamma \eta}{(k+\sigma\gamma)(k+\sigma\eta)}\big) \Big)$, when δ_∞ is large enough, more precisely when

$$\delta_\infty \geq \Delta_\infty = C_\infty + \frac{\sigma}{k} (2C_2 + C_1 \bar{q}^2) T, \tag{19}$$

see Proposition 1 in Chapter IX. Furthermore, the optimal control $\hat{z}(t,q) = (\hat{z}^s(t,q), \hat{z}^a(t,q), \hat{z}^b(t,q))$ is given by

$$\hat{z}^s(t,q) = -\frac{\gamma}{\gamma+\eta} q, \; \hat{z}^a(t,q) = \zeta_0 + \frac{1}{\eta} \log\Big(\frac{v(t,q)}{v(t,q-1)} \Big), \text{ and } \hat{z}^b(t,q) = \zeta_0 + \frac{1}{\eta} \log\Big(\frac{v(t,q)}{v(t,q+1)} \Big).$$

with $\zeta_0 = c + \frac{1}{\eta} \log\Big(1 - \frac{\sigma^2 \gamma \eta}{(k+\sigma\gamma)(k+\sigma\eta)} \Big)$.

Note that, applying the change of variable $u = (-v)^{-\frac{k}{\sigma\eta}}$, this HJB equation (18) is reduced to a linear one

$$\begin{cases} \partial_t u(t,q) - C_1 q^2 u(t,q) + C_2 \big(u(t,q+1) \mathbb{1}_{\{q<\bar{q}\}} + u(t,q-1) \mathbb{1}_{\{q>-\bar{q}\}} \big) = 0, \quad t \in [0,T), \\ u(T,q) = 1, \end{cases}$$

leading to the existence of a unique solution to (18).

Result 20. *Assume that δ_∞ satisfies (19), then the optimal contract for the problem of the exchange (15) is given by*

$$\hat{\xi} \;=\; \hat{Y}_0 + \int_0^T \hat{Z}_r^a dN_r^a + \hat{Z}_r^b dN_r^b + \hat{Z}_r^S dS_r + \Big(\frac{1}{2} \gamma \sigma^2 \big(\hat{Z}_r^S + Q_r \big)^2 - H\big(\hat{Z}_r, Q_r \big) \Big) dr,$$

with $\hat{Z}_r^S = \hat{z}^s(r, Q_{r-})$, $\hat{Z}_r^a = \hat{z}^a(r, Q_{r-})$, and $\hat{Z}_r^b = \hat{z}^b(r, Q_{r-})$. The market maker's optimal effort is given by

$$\hat{\delta}_t^a = \hat{\delta}_t^a(\hat{\xi}) = -\hat{Z}_t^a + \frac{1}{\gamma} \log(1 + \frac{\sigma\gamma}{k}), \quad \hat{\delta}_t^b = \hat{\delta}_t^b(\hat{\xi}) = -\hat{Z}_t^b + \frac{1}{\gamma} \log(1 + \frac{\sigma\gamma}{k}).$$

The optimal contract $\hat{\xi}$ depends on the inventory trajectory of the market maker. More precisely, we notice that $\hat{Z}^a$ (resp. $\hat{Z}^b$) is an increasing (resp. decreasing) function of Q. Hence, when the inventory is highly positive (resp. negative), the exchange is encouraging the market maker to attract buy (resp. sell) market orders. Furthermore, the presence of the integral $\int_0^T \hat{Z}_r^S dS_r$ means that the exchange shares the inventory risk of the market maker with the rate $\frac{\gamma}{\gamma+\eta}$. This risk sharing has the role to encourage the market maker to take more inventory risk by accepting more market orders.

From Result 20, the optimal spread is given

$$-2c + \frac{\sigma}{k}\log\Big(\frac{u(t,Q_{t-})^2}{u(t,Q_{t-}-1)u(t,Q_{t-}+1)}\Big) - \frac{2}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big) + \frac{2}{\gamma}\log(1+\frac{\sigma\gamma}{k}).$$

In practice,

$$\frac{u(t,q)^2}{u(t,q-1)u(t,q+1)}$$

is close to unity and $\sigma\gamma/k$ is small. Thus, the optimal spread is approximated by

$$-2c + \frac{2\sigma}{k}.$$

Remarking that the optimal utility V_0^E does not depend on the transaction cost c, the exchange may fix c so that the spread is fixed to its minimal value of one tick by taking

$$c \approx \frac{\sigma}{k} - \frac{1}{2}\text{Tick}.$$

7.4 Impact of the optimal policy

In Section 5 of Chapter IX, we compare the results obtained under the optimal policy with the situation without incentive policy from the exchange towards the market maker activities, that is the case $\xi = 0$, considered in [AS08, GLFT13]. In particular, we obtain that

- The incentive policy reduces the spread and hence increases the liquidity of the market and the average number of market orders occurred in the platform.
- The incentive policy increases the expected total profit and loss of the market maker and the exchange.
- The incentive policy reduces the total transaction costs paid by the market taker.

Part I

The microstructural foundations of rough volatility

CHAPTER I

Microstructural foundations of leverage effect and rough volatility

Abstract

We show that typical behaviors of market participants at the high frequency scale generate leverage effect and rough volatility. To do so, we build a simple microscopic model for the price of an asset based on Hawkes processes. We encode in this model some of the main features of market microstructure in the context of high frequency trading: high degree of endogeneity of market, no-arbitrage property, buying/selling asymmetry and presence of metaorders. We prove that when the first three of these stylized facts are considered within the framework of our microscopic model, it behaves in the long run as a Heston stochastic volatility model, where leverage effect is generated. Adding the last property enables us to obtain a rough Heston model in the limit, exhibiting both leverage effect and rough volatility. Hence we show that at least part of the foundations of leverage effect and rough volatility can be found in the microstructure of the asset.

Keywords: Market microstructure, high frequency trading, leverage effect, rough volatility, Hawkes processes, limit theorems, Heston model, rough Heston model.

1 Introduction

Leverage effect is a well-known stylized fact of financial data. It refers to the negative correlation between price returns and volatility increments: when the price of an asset is increasing, its volatility drops, while when it decreases, the volatility tends to become larger. The name "leverage" comes from the following interpretation of this phenomenon due to Black [Bla76] and Christie [Chr82]: When an asset price declines, the associated company becomes automatically more leveraged since the ratio of its debt with respect to the equity value becomes larger. Hence the risk of the asset, namely its volatility, should become more important. Another economic interpretation of the leverage effect, inverting causality, is that the forecast of an increase of the volatility should be compensated by a higher rate of return, which can only be obtained through a decrease in the asset value, see [CH92, FW00, FSS87].

From an empirical viewpoint, leverage effect and the plausible interpretations for it have been widely studied in the literature, see for example [BW00, BLT06, EN93, WX02]. Furthermore, some statistical methods enabling us to use high frequency data have been built to measure it, see [ASFL13, WM14]. From a modeling perspective, the will to reproduce the leverage phenomenon has been a key motivation in the development of sophisticated time series models, for example of ARCH type, see [BCK92, DGE93, Nel91, RR12, Zak94]. Finally, in financial engineering, it has become clear in the late eighties that it is necessary to introduce leverage effect in derivatives pricing frameworks in order to accurately reproduce the behavior of the implied volatility surface. This led to the rise of famous stochastic volatility models, where the Brownian motion driving the volatility is (negatively) correlated with that driving the price, see for example [HKLW02, Hes93, HW87, SS91] for SABR, Heston, Hull and White and Stein and Stein stochastic volatility models.

As mentioned above, traditional explanations for leverage effect are based on “macroscopic" arguments from financial economics. In this paper, we wish to address the following question: Could microscopic interactions between agents naturally lead to leverage effect at larger time scales? Hence we would like to know whether part of the foundations for leverage effect could be microstructural. To do so, our idea is to consider a very simple agent-based model, encoding well-documented and understood behaviors of market participants at the microscopic scale. Then we aim at showing that in the long run, this model leads to a price dynamic exhibiting leverage effect. This would demonstrate that typical strategies of market participants at the high frequency level naturally induce leverage effect.

One could argue that transactions take place at the finest frequencies and prices are revealed through order book type mechanisms. Therefore, it is an obvious fact that leverage effect arises from high frequency properties. However, what we wish to show here is that under certain market conditions, typical high frequency behaviors, having probably no connection with the financial economics concepts mentioned earlier, may give rise to some leverage effect at the low frequency scales. It is important to emphasize that we do not claim that leverage effect should be fully explained by high frequency features. What we simply say is that part of it could be generated from the microstructure of the asset.

Another important stylized fact of financial data, which has been highlighted recently in [GJR18], is the rough nature of the volatility process. Indeed, it is shown in [GJR18] that for a very wide range of assets, historical volatility time-series exhibit a behavior which is much rougher than that of a Brownian motion. More precisely, the dynamics of the log-volatility are typically very well modeled by a fractional Brownian motion with Hurst parameter around 0.1, that is a process with Hölder regularity of order 0.1. Furthermore, using a fractional Brownian motion with small Hurst index also enables us to reproduce very accurately the features of the volatility surface, see [BFG16, GJR18].

The fact that for basically all reasonably liquid assets, volatility is rough, with the same order of magnitude for the roughness parameter, is of course very intriguing. Thus we also aim in this work at understanding how such a surprising feature can be generated. Some elements in this direction are already provided in [JR16b]. Here we want to go further and investigate the behavior of the long term volatility in our microscopic model encoding the main stylized facts of modern market microstructure. We wish to show that the rough nature of the volatility naturally emerges from typical behaviors of market participants at the high frequency scale.

Our tick-by-tick price model is based on a bi-dimensional Hawkes process, very much inspired by the approaches in [BDHM13a, BDHM13b, JR15]. A bi-dimensional Hawkes process is a bivariate point process $(N_t^+, N_t^-)_{t\geq 0}$ taking values in $(\mathbb{R}_+)^2$ and with intensity $(\lambda_t^+, \lambda_t^-)$ of the form

$$\begin{pmatrix}\lambda_t^+\\ \lambda_t^-\end{pmatrix} = \begin{pmatrix}\mu^+\\ \mu^-\end{pmatrix} + \int_0^t \begin{pmatrix}\varphi_1(t-s) & \varphi_3(t-s)\\ \varphi_2(t-s) & \varphi_4(t-s)\end{pmatrix}.\begin{pmatrix}dN_s^+\\ dN_s^-\end{pmatrix}.$$

Here μ^+ and μ^- are positive constants and the functions $(\varphi_i)_{i=1,\dots 4}$ are non-negative with associated matrix called kernel matrix, see Section 2.1 for further details. Hawkes processes have been introduced by Hawkes in [Haw71]. They are said to be self-exciting, in the sense that the instantaneous jump probability depends on the location of the past events. Hawkes processes are nowadays of standard use in finance, not only in the field of microstructure but also in risk management or contagion modeling, see among many others [ASCDL15, BDHM13a, BH04, Bow07, CDDM05, ELL11, EGG10, JR15, JR16b]. It is explained in [BDHM13a] that a relevant model for the ultra high frequency dynamic of the price P_t of a large tick asset[1] is simply given by

$$P_t = N_t^+ - N_t^-.$$

Thus, in this approach, N_t^+ corresponds to the number of upward jumps of the asset in the time interval $[0,t]$ and N_t^- to the number of downward jumps. Hence, the instantaneous probability to get an upward (downward) jump depends on the arrival times of the past upward and downward jumps. Furthermore, by construction, the price process lives on a discrete grid, which is obviously a crucial feature of high frequency prices in practice. Statistical properties of this model have been studied in details in [BDHM13a]. In particular, it is shown that such dynamic is very convenient in order to reproduce the commonly observed bid-ask bounce effect.

This simple tick-by-tick price model enables us to encode very easily the following important stylized facts of modern electronic markets in the context of high frequency trading:

i) Markets are highly endogenous, meaning that most of the orders have no real economic motivation but are rather sent by algorithms in reaction to other orders, see [FS15, HBB13] and Section 2.1.3 for more details.

ii) Mechanisms preventing statistical arbitrages take place on high frequency markets. Indeed, at the high frequency scale, building strategies which are on average profitable is hardly possible, see [ALR14].

[1] A large tick asset is an asset whose bid-ask spread is almost always equal to one tick and therefore essentially moves by one tick jumps, see [DR15].

iii) There is some asymmetry in the liquidity on the bid and ask sides of the order book. This simply means that buying and selling are not symmetric actions. Indeed, consider for example a market maker, with an inventory which is typically positive. He is likely to raise the price by less following a buy order than to lower the price following the same size sell order. This is because its inventory becomes smaller after a buy order, which is a good thing for him, whereas it increases after a sell order, see [BCST12, BP09, HS06, HS81, TT12].

iv) A significant proportion of transactions is due to large orders, called metaorders, which are not executed at once but split in time by trading algorithms, see [AC01, LL13].

In a Hawkes process framework, the first of these properties corresponds to the case of so-called ***nearly unstable Hawkes processes***, that is Hawkes processes for which the stability condition is almost saturated. This means the spectral radius of the kernel matrix integral is smaller than but close to unity, see [FS15, HBB13, JR15, JR16b]. The second and third ones impose a specific structure on the kernel matrix and the fourth one leads to functions φ_i with heavy tails, see [JR16b]. The parametrization of our price process corresponding to the four properties above is developed in more details in Sections 2.1 and 3.1.

In this work, we study the long term behavior of such Hawkes-based ultra high frequency price models, for which the parameters are consistent with the four mentioned properties of market microstructure. Doing so, we investigate the macroscopic price dynamics arising from a situation where the four ingredients above are put together. More precisely, we start with the case of a Hawkes-based model where Properties i, ii and iii only are satisfied. Our first result states that in this setting, the macroscopic dynamic of the price is that of a Heston stochastic volatility model as introduced in [Hes93], where the volatility is (negatively) correlated with the price. Hence leverage effect is produced. This extends some results in [JR15] where a non-correlated Heston limit is obtained. Then, when in addition Property iv is encoded in our microscopic model, we show that a so-called rough-Heston model, where the volatility is rough and negatively correlated with the price, is generated at low frequency. More precisely, as in [GJR18], the volatility process is driven by a fractional Brownian motion with Hurst parameter smaller than 1/2.

The computations and techniques applied in this paper are partially inspired by the papers [JR15, JR16b]. In [JR15, JR16b], a rigorous connection is established between one-dimensional nearly unstable Hawkes processes and (fractional) Cox-Ingersoll-Ross processes. This relation aimed at linking the behavior of the order flow to that of the integrated variance. Here we do not focus on the order flow and integrated variance but on a link between the microscopic price process and the macroscopic price dynamic. This requires to work in a more intricate multidimensional setting but enables us to derive more satisfactory microstructural foundations for the low frequency behavior of prices. Indeed, in practice, one observes prices and not volatility. Of course our results are not the first ones relating high frequency dynamics to long term behaviors with stochastic volatility. The most famous example is probably that of Nelson who shows in [Nel90] that in specific settings, GARCH processes converge to (uncorrelated) stochastic volatility models, see also [Cor00, Dua97, Lin09]. However, to our knowledge, we provide the first natural, non ad-hoc approach allowing for leverage effect, and even rough volatility, in the long term limit of the price dynamic.

The paper is organized as follows. In Section 2, we parametrize our Hawkes-based microscopic price model so that Properties i, ii and iii are satisfied. Then we show that after proper rescaling, this price converges in the long run to a Heston stochastic volatility model where leverage effect is observed. In Section 3, we incorporate Property iv into our microscopic model and prove that it leads to a rough Heston model at the macroscopic scale, where leverage effect is still generated. Some proofs are relegated to Section 4 and some useful technical results are given in an appendix.

Notations Let $\mathbb{R}_+$ and $\mathbb{R}_+^*$ be respectively the set of real nonnegative and positive numbers and $\mathbb{N}^* = \mathbb{N} - \{0\}$ be the set of positive integers. For any $(n,m) \in (\mathbb{N}^*)^2$, $\mathcal{M}^{n,m}(\mathbb{R})$ denotes the space of $n \times m$ matrices with values in $\mathbb{R}$ and $\mathcal{M}^n(\mathbb{R})$ denotes $\mathcal{M}^{n,n}(\mathbb{R})$. For $d \in \mathbb{N}^*$, the space of d-dimensional vectors with real values $\mathbb{R}^d$ will be identified with $\mathcal{M}^{d,1}(\mathbb{R})$. For $X \in \mathcal{M}^{l,m}(\mathbb{R})$ and $Y \in \mathcal{M}^{m,n}(\mathbb{R})$, $X.Y$ stands for the usual matrix multiplication and $X^\top$ refers to the transpose of X. Finally we denote by $\cdot$ for the inner production, namely $u \cdot v = u^\top.v$ for any $u, v \in \mathbb{R}^n$.

2 From high frequency features to leverage effect

We build in this section a Hawkes-based microscopic tick-by-tick model in which Properties i, ii and iii are satisfied. This leads us to a specific parametrization of our Hawkes process. We show that after suitable rescaling, the long term price dynamic becomes that of a Heston model. We start by defining our microscopic price model.

2.1 Building a suitable microscopic price model

2.1.1 The Hawkes process framework

We consider a tick-by-tick price model based on a bi-dimensional Hawkes process $N_t = (N_t^+, N_t^-)$, with intensity $\lambda_t = (\lambda_t^+, \lambda_t^-)$ defined by

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \varphi_1(t-s) & \varphi_3(t-s) \\ \varphi_2(t-s) & \varphi_4(t-s) \end{pmatrix} \cdot \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix},$$

where μ^+ and μ^- are positive constants and

$$\phi = \begin{pmatrix} \varphi_1 & \varphi_3 \\ \varphi_2 & \varphi_4 \end{pmatrix} : \mathbb{R}_+ \to \mathcal{M}^2(\mathbb{R}_+^*)$$

is a kernel matrix whose components φ_i are positive and locally integrable. Inspired by [BDHM13a, BDHM13b, JR15], our model for the ultra high frequency transaction price P_t is simply given by

$$P_t = N_t^+ - N_t^-.$$

Thus N_t^+ is the number of upward jumps of one tick of the asset in the time interval $[0,t]$ and N_t^- is the number of downward jumps of one tick of the asset in the time interval $[0,t]$.

Let us now interpret the intensity process λ_t^+ (interpretation for λ_t^- goes similarly). At time t, the probability to get a new one-tick upward jump between t and $t+dt$ is given by $\lambda_t^+ dt$. This probability can be decomposed into three terms:

- $\mu_+ dt$, which is the Poissonian part of the intensity and therefore corresponds to the probability that the price goes up because of some exogenous reason.
- $\left(\int_0^t \varphi_1(t-s)dN_s^+\right)dt$, which is the probability of upward jump induced by past upward jumps.
- $\left(\int_0^t \varphi_3(t-s)dN_s^-\right)dt$, which is the probability of upward jump induced by past downward jumps.

In particular, we see here that when the φ_i have suitable shapes, it is easy to reproduce the bid-ask bounce effect by imposing a high probability of upward (resp. downward) jump right after a downward (resp. upward) jump.

2.1.2 Encoding Properties ii and iii

We now provide a specific structure on the parameters of the intensity process so that Properties ii and iii are satisfied in our model. Property ii is the no-statistical arbitrage condition. In a high frequency setting, this amounts to say that on average, there should be essentially as many upward as downward jumps on any given time-period. We translate this within our Hawkes framework noting that

$$\mathbb{E}[N_t^+] = \int_0^t \mathbb{E}[\lambda_s^+]ds, \quad \mathbb{E}[N_t^-] = \int_0^t \mathbb{E}[\lambda_s^-]ds,$$

and

$$\begin{aligned}
\mathbb{E}[\lambda_t^+] &= \mu^+ + \int_0^t \varphi_1(t-s)\mathbb{E}[\lambda_s^+]ds + \int_0^t \varphi_3(t-s)\mathbb{E}[\lambda_s^-]ds, \\
\mathbb{E}[\lambda_t^-] &= \mu^- + \int_0^t \varphi_2(t-s)\mathbb{E}[\lambda_s^+]ds + \int_0^t \varphi_4(t-s)\mathbb{E}[\lambda_s^-]ds.
\end{aligned}$$

Therefore we obtain that a simple and natural way to implement the no-statistical arbitrage condition is to set $\mathbb{E}[\lambda_t^+] = \mathbb{E}[\lambda_t^-]$ by imposing

$$\mu^+ = \mu^- \quad \text{and} \quad \varphi_1 + \varphi_3 = \varphi_2 + \varphi_4.$$

In terms of microscopic price movements, Property iii, which states that the ask side is more liquid than the bid side, can be translated as follows: the conditional probability to observe an upward jump right after an upward jump is smaller than the conditional probability to observe a downward jump right after a downward jump. In our Hawkes framework, it amounts to have $\varphi_1(x) < \varphi_4(x)$ or similarly $\varphi_3(x) > \varphi_2(x)$ when x is close to zero. For simplicity and technical convenience, we in fact make the more restrictive assumption that there exists some $\beta > 1$ such that

$$\varphi_3 = \beta\varphi_2.$$

Therefore we assume the following structure for the intensity process

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \mu \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \int_0^t \phi(t-s).\begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix}, \tag{1}$$

where

$$\phi = \begin{pmatrix} \varphi_1 & \beta\varphi_2 \\ \varphi_2 & \varphi_1 + (\beta-1)\varphi_2 \end{pmatrix},$$

with $\mu > 0$ and $\beta \geq 1$. We now explain how to deal with Property i.

2.1.3 Dealing with Property i: Nearly unstable Hawkes processes

Property i states that modern markets are highly endogenous. To understand how this high degree of endogeneity can be translated through our Hawkes-based price model, let us consider for simplicity a one-dimensional Hawkes process $\tilde{N}_t$ with intensity

$$\tilde{\lambda}_t = \tilde{\mu} + \int_0^t \tilde{\varphi}(t-s)d\tilde{N}_s,$$

where $\tilde{\mu} > 0$ and $\tilde{\varphi}$ is a non-negative measurable function such that its L^1 norm $||\tilde{\varphi}||_1$ satisfies $||\tilde{\varphi}||_1 < 1$. This last constraint is called stability condition and plays the same role as that which states that the coefficient of an order 1 auto-regressive process has to be smaller than one, see [JR15]. In particular, this condition ensures the existence of a stationary solution for the intensity (when time starts at $-\infty$). Such one-dimensional Hawkes processes are usually considered to model order flows, see [BMM15] and the references therein. So $\tilde{N}_t$ can typically be viewed as the number of transactions in the time interval $[0, t]$.

From a probabilistic viewpoint, the cluster representation of Hawkes processes, see [HO74], enables to see $\tilde{N}$ as a population process. In this population, migrants arrive following a Poisson process with intensity $\tilde{\mu}$. Each migrant gives birth to children according to an inhomogeneous Poisson process with intensity $\tilde{\varphi}$. Then each child also gives birth to children according to an inhomogeneous Poisson process with intensity $\tilde{\varphi}$ and so on. Coming back to financial markets, let us consider a dichotomy between “economic" (or exogenous) orders, which are executed because some market participants have a fundamental will to buy or sell, and endogenous orders, which are just sent in reaction to other orders. In the Hawkes context, it is therefore very natural to make the following interpretation: exogenous orders correspond to migrants and endogenous orders to descendants of migrants, see [FS15, HBB13, JR15].

Now remark that each migrant, or descendant of a migrant, has on average $\|\tilde{\varphi}\|_1$ children. Hence a migrant has on average

$$\sum_{k\geq 1}\|\tilde{\varphi}\|_1^k = \frac{\|\tilde{\varphi}\|_1}{1-\|\tilde{\varphi}\|_1}$$

descendants. Now, the number of people in a "family" being the number of descendants plus one (the plus one corresponding to the initial migrant), the proportion of descendants in the whole population is given by $||\tilde{\varphi}||_1$. In our financial interpretation, it means that $||\tilde{\varphi}||_1$ corresponds to the proportion of endogenous orders in the market. Hence, to get a model which is in agreement with Property i, we need to take $||\tilde{\varphi}||_1$ smaller than but close to unity. This situation is called nearly unstable case, and is actually in agreement with the empirical measurements for $||\tilde{\varphi}||_1$ made in [BDM12, FS15, HBB13].

Let us now come back to our bi-dimensional Hawkes process of interest with intensity defined by (1). In the same way as in the one-dimensional case, one can define the degree of endogeneity as the spectral radius of the kernel matrix integral, that is

$$\mathscr{S}(\int_0^\infty \phi(s)ds) = \|\varphi_1\|_1 + \beta\|\varphi_2\|_1,$$

where $\mathscr{S}$ denotes the spectral radius operator. We want to assume that this spectral radius is smaller than but close to unity. To do so, we introduce an asymptotic framework, in the spirit of [JR15, JR16b]. More precisely, we work on a sequence of probability spaces $(\Omega^T, \mathscr{F}^T, \mathbb{P}^T)$, indexed by $T>0$, on which $N^T = (N^{T,+}, N^{T,-})$ is a bi-dimensional Hawkes process defined on $[0,T]$ and with intensity of the form

$$\lambda_t^T = \begin{pmatrix}\lambda_t^{T,+}\\ \lambda_t^{T,-}\end{pmatrix} = \mu_T\begin{pmatrix}1\\1\end{pmatrix} + \int_0^t \phi^T(t-s).dN_s^T. \tag{2}$$

For given T, the probability space is equipped with the filtration $(\mathscr{F}_t^T)_{t\geq 0}$, where $\mathscr{F}_t^T$ is the σ-algebra generated by $(N_s^T)_{s\leq t}$. Respecting the constraints on the parameters given in (1) and taking into account the discussion above about the endogeneity of the market, we make the following assumption on λ_t^T.

Assumption 1. *We have $\mu_T > 0$ and*

$$\phi^T = a_T\phi,\quad \phi = \begin{pmatrix}\varphi_1 & \beta\varphi_2\\ \varphi_2 & \varphi_1 + (\beta-1)\varphi_2\end{pmatrix},$$

where $\beta\geq 1$, φ_1 and φ_2 are two positive measurable functions such that

$$\mathscr{S}(\int_0^\infty \phi(s)ds) = \|\varphi_1\|_1 + \beta\|\varphi_2\|_1 = 1$$

and a_T is an increasing sequence of positive numbers converging to one.

From now on, our microscopic price process is given by

$$P_t^T = N^{T,+} - N^{T,-}.$$

Thus, under Assumption 1, we are indeed working in the nearly unstable case since

$$\mathscr{S}(\int_0^\infty \phi^T(s)ds) = a_T.$$

Therefore, our microscopic price process P^T reproduces Properties i, ii and iii. We now focus on the asymptotic behavior of P^T.

2.2 The macroscopic limit of the high frequency model with leverage

We give in this section our convergence result for the microscopic price towards a Heston model. In fact such result can be found in [JR15] in the case $\beta = 1$. As in [JR15], we need to consider the following assumption on the asymptotic framework and the kernel function.

Assumption 2. *There exist positive parameters λ, μ and m such that*

$$T(1-a_T) \underset{T\to\infty}{\to} \lambda, \quad \mu_T = \mu,$$

and

$$\mathscr{S}(\int_0^\infty x\phi(x)dx) = m < \infty.$$

It is well-explained in [JR15] that in the light-tailed case (which is characterized by the second part of Assumption 2), only one asymptotic setting enables us to recover a non-degenerate limit. This is when the kernel L^1 norm a_T goes to unity such that $T(1-a_T)$ is of order one. Now let

$$\psi^T = \sum_{k\geq 1} (\phi^T)^{*k},$$

where $(\phi^T)^{*1} = \phi^T$ and for $k > 1$, $(\phi^T)^{*k}(t) = \int_0^t \phi^T(s)(\phi^T)^{*(k-1)}(t-s)ds$. The following technical assumption is also required in [JR15].

Assumption 3. *The function $\psi^T = \sum_{k\geq 1}(\phi^T)^{*k}$ is uniformly bounded and ϕ is differentiable such that each component ϕ_{ij} satisfies $||\phi'_{ij}||_\infty < \infty$ and $||\phi'_{ij}||_1 < \infty$.*

The uniform boundedness assumption here is not really restrictive. Indeed, as it will be clear from the computations in Section 2.3.2, a sufficient condition for it is the fact that the largest eigenvalue of ϕ is non-increasing, see also [JR15]. In our model, this is for example the case if both φ_1 and φ_2 are non-increasing.

When $\beta = 1$, under Assumptions 1, 2 and 3, it is proved in [JR15] that the rescaled price process

$$\frac{1}{T}P_{tT}^T = \frac{N_{tT}^{T,+} - N_{tT}^{T,-}}{T}$$

converges in law over $[0,1]$ towards a Heston model defined by

$$P_t = \frac{1}{1-(\|\varphi_1\|_1 - \|\varphi_2\|_1)} \int_0^t \sqrt{X_s} dW_s,$$

with

$$dX_t = \frac{\lambda}{m}(\frac{2\mu}{\lambda} - X_t)dt + \frac{1}{m}\sqrt{X_t} dB_t, \quad X_0 = 0,$$

where W and B are two independent Brownian motions. Remark that in this setting, the time scale T is at the same time the reciprocal of price tick size. Thus as T goes to infinity, the price moves more frequently with smaller size. Such an asymptotic analysis enables us to derive a stochastic volatility process as a macroscopic model. Note also that $\mu_T = \mu > 0$ in Assumption 2 is quite intuitive in this scaling, meaning that exogenous moves of size $1/T$ occur μT times on average in a unit time interval, maintaining non-zero contribution from exogenous activity in the limit.

However, when $\beta = 1$, the important Property iii about the liquidity asymmetry between the bid and ask sides of the order book is not reproduced in the dynamic of the microscopic price. Our first main theorem below shows that this property, encoded by the fact that $\beta > 1$, is the microscopic feature at the origin of leverage effect at low frequency.

Theorem 1. *Under Assumptions 1, 2 and 3, as T tends to infinity, the rescaled microscopic price*

$$\frac{1}{T} P^T_{tT} = \frac{N^{T,+}_{tT} - N^{T,-}_{tT}}{T}, \quad t \in [0,1],$$

converges in law for the Skorokhod topology to the following Heston model[2]

$$P_t = \frac{1}{1-(\|\varphi_1\|_1 - \|\varphi_2\|_1)} \sqrt{\frac{2}{1+\beta}} \int_0^t \sqrt{X_s} dW_s,$$

with

$$dX_t = \frac{\lambda}{m}\big((\beta+1)\frac{\mu}{\lambda} - X_t\big)dt + \frac{1}{m}\sqrt{\frac{1+\beta^2}{1+\beta}} \sqrt{X_t} dB_t, \quad X_0 = 0,$$

where (W, B) is a correlated bi-dimensional Brownian motion with

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}} dt.$$

Hence, putting Properties i, ii and iii together in a simple but reasonable way (through the microscopic price P^T), we naturally obtain stochastic volatility and leverage effect in the long run. Indeed, when $\beta > 1$, the asymmetry in the liquidity at the microstructural level

[2] Notice that the limit price is actually a "Bachelier" version of the Heston model. Furthermore, remark that the definition of the rough Heston model is not a well-established one and other types of fractional Heston models can be defined, see for example [GJR14].

generates a negative correlation between low frequency price returns and volatility increments. Nevertheless, Properties i and ii are also crucial in order to obtain Theorem 1. In fact, no stochastic volatility can be obtained without Property i and the failure of Property ii would lead to a drift process in the limit.

Finally, note that from a technical point of view, to our knowledge, this result is the first scaling limit of a microscopic price process inducing leverage effect in the long run in a non ad-hoc way.

We now give in the next section a general result about the convergence of nearly unstable multidimensional Hawkes processes. This result is the key element of the proof of Theorem 1.

2.3 Convergence of nearly unstable multidimensional Hawkes processes

2.3.1 Setting

In order to show Theorem 1, we study the convergence of a general sequence of nearly unstable d-dimensional Hawkes processes defined on $[0,T]$, with T tending to infinity and $d \in \mathbb{N}^*$. We keep the notation N^T for our Hawkes process of interest whose intensity λ^T is defined by

$$\lambda_t^T = \mu_T \mathbf{1} + \int_0^t \phi^T(t-s).dN_s^T,$$

where $\mu_T > 0$ and $\phi^T = a_T \phi$, with a_T an increasing sequence of positive numbers converging to unity, and the matrix $\phi : \mathbb{R}_+ \to \mathscr{M}^d(\mathbb{R}_+^*)$ has integrable components such that

$$\mathscr{S}(\int_0^\infty \phi(s)ds) = 1.$$

We use also the notation

$$M_t^T = N_t^T - \int_0^t \lambda_s^T ds$$

for the martingale associated to N^T. We furthermore assume that for any $t \geq 0$, $\phi(t)$ is diagonalizable on $\mathbb{R}$. We write $\lambda_1(t) \geq \cdots \geq \lambda_d(t)$ for the eigenvalues of $\phi(t)^\top$ and $v_1, \ldots, v_d \in \mathbb{R}^d$ for the corresponding eigenvectors. We assume that these eigenvectors do not depend on t (as it is the case under Assumption 1). We also recall that from Frobenius-Perron theorem, for $i \geq 2$, $|\lambda_i(t)| < \lambda_1(t) = \mathscr{S}\big(\phi(t)\big)$ and v_1 can be taken in $\mathbb{R}_+^d$. Notice that in this setting, $\int_0^\infty \lambda_1(s)ds = \int_0^\infty \mathscr{S}(\phi(s))ds = \mathscr{S}(\int_0^\infty \phi(s)ds) = 1$ and that for $i \geq 2$, $\|\lambda_i\|_1 < 1$.

Finally, we define an orthonormal basis $(e_1, \ldots, e_d)$ of $\mathbb{R}^d$ such that $e_1 \cdot v_1 > 0$ and

$$\text{span}(e_2, \ldots, e_d) = \text{span}(v_2, \ldots, v_d)$$

and set $v' = e_1 - \frac{1}{e_1 \cdot v_1} v_1$. Note that v' belongs to $\text{span}(v_2, \ldots, v_d)$.

2.3.2 Intuition for the result and theorem

We now provide some non-rigorous developments which are helpful in order to understand the asymptotic behavior of the multidimensional process N^T. We work under Assumptions 2 and 3 and the setting of Section 2.3.1. We have

$$\lambda_t^T = \mu_T \mathbf{1} + \int_0^t \phi^T(t-s).dM_s^T + \int_0^t \phi^T(t-s).\lambda_s^T ds.$$

Using Lemma 2 in Appendix together with the Fubini theorem and the fact that the convolution product $\psi^T * \phi^T$ satisfies $\psi^T * \phi^T = \psi^T - \phi^T$, we get

$$\lambda_t^T = \mu_T \mathbf{1} + \mu_T \int_0^t \psi^T(t-s)ds.\mathbf{1} + \int_0^t \psi^T(t-s).dM_s^T, \tag{3}$$

where

$$\psi^T = \sum_{k\geq 1} (\phi^T)^{*k} = \sum_{k\geq 1} a_T^k \phi^{*k}.$$

The function ψ^T being uniformly bounded and μ_T being constant equal to $\mu > 0$, we get by the same procedure used in [JR15], that $\mathbb{E}[\lambda_{tT}^T]$ is of order T. Thus a natural rescaling in time and space leads us to consider for $t \in [0,1]$ the process

$$C_t^T = \frac{1}{T}\lambda_{tT}^T.$$

From (3), we obtain

$$C_t^T = \frac{\mu}{T}\mathbf{1} + \mu \int_0^t \psi^T\big(T(t-s)\big)ds.\mathbf{1} + \int_0^t \psi^T\big(T(t-s)\big).d\overline{M}_s^T,$$

with $\overline{M}_t^T = M_{tT}^T / T$. Note that since

$$\langle M^T, M^T\rangle_t = \mathrm{diag}(\int_0^t \lambda_s^T ds),$$

we get that

$$\mathbb{E}[\langle \overline{M}^T, \overline{M}^T\rangle_t] = \frac{1}{T^2}\mathbb{E}\big[\mathrm{diag}(\int_0^{tT} \lambda_s^T ds)\big] = \mathrm{diag}(\int_0^t \mathbb{E}[C_s^T]ds)$$

is bounded. Now remark that for each $i \in \{1,\dots,d\}$, using a recursion, we easily see that for any $k \geq 1$, $v_i^\top.\phi^{*k}(t) = \lambda_i^{*k}(t) v_i^\top$. Consequently, defining for $i \in \{1,\dots,d\}$

$$\psi_i^T = \sum_{k\geq 1} a_T^k \lambda_i^{*k},$$

we have

$$v_i^\top.\psi^T = \psi_i^T v_i^\top.$$

Hence we can write the dynamic of $v_i \cdot C_t^T$ as follows

$$v_i \cdot C_t^T = \frac{\mu}{T}(v_i \cdot \mathbf{1}) + \mu (v_i \cdot \mathbf{1}) \int_0^t \psi_i^T\big(T(t-s)\big)ds + \int_0^t \psi_i^T\big(T(t-s)\big)(v_i \cdot d\overline{M}_s^T). \tag{4}$$

Thus, to understand the asymptotic behavior of $v_i \cdot C^T$ as T goes to infinity, we need to study that of the functions $\psi_i^T(T.)$. To do so, one can compute the Fourier transform $\hat{\psi}_j^T(T.)$ of $\psi_j^T(T.)$ for each $j \in \{1,\dots,d\}$. We have

$$\hat{\psi}_j^T(T.)(z) = \int_{x\in\mathbb{R}_+} \psi_j^T(Tx)e^{ixz}dx = \frac{1}{T}\sum_{k\geq 1} a_T^k\big(\hat{\lambda}_j(z/T)\big)^k = \frac{a_T\hat{\lambda}_j(z/T)}{T\big(1-a_T\hat{\lambda}_j(z/T)\big)}.$$

Now, as T goes to infinity, $\hat{\lambda}_j(z/T)$ tends to $\|\lambda_j\|_1$ and recall that $\|\lambda_j\|_1 < 1$ for $j \geq 2$. Thus, for $j \geq 2$, $\psi_j^T(T.)$ should asymptotically vanish, as should consequently be the case for $v_j \cdot C^T$.

For $j = 1$, using Assumption 2 and the same approach developed in [JR15], we obtain that

$$\hat{\psi}_1^T(T.)(z) \underset{T\to\infty}{\to} \frac{1}{\lambda - izm},$$

which is the Fourier transform of

$$x \in \mathbb{R}_+ \to \frac{1}{m}e^{-\frac{\lambda}{m}x}.$$

Hence we can expect that $\psi_1(Tx)$ converges to $\frac{1}{m}e^{-\frac{\lambda}{m}x}$.

Let us now deduce from the preceding computation the behavior of $v_1 \cdot C^T$. From (4), this quantity can be written

$$v_1 \cdot C_t^T = \frac{\mu}{T}(v_1\cdot\mathbf{1}) + \mu(v_1\cdot\mathbf{1})\int_0^t \psi_1^T\big(T(t-s)\big)ds + \int_0^t \psi_1^T\big(T(t-s)\big)\sqrt{v_1^2\cdot C_s^T}\,dB_s^T, \tag{5}$$

where $v_1^2 = (v_{1,i}^2)_{1\leq i\leq d}$ and

$$B_t^T = \int_0^{tT} \frac{v_1\cdot dM_s^T}{\sqrt{Tv_1^2\cdot\lambda_s^T}}. \tag{6}$$

The sequence of processes B^T has been specifically chosen since the associated sequence of quadratic variations converges to identity. Thus the limit of B^T is a Brownian motion.

Decomposing v_1^2 in the basis $(e_1,\dots,e_d)$ defined in Section 2.3.1, we get

$$v_1^2\cdot C_t^T = \frac{e_1\cdot v_1^2}{e_1\cdot v_1}(v_1\cdot C_t^T) + (e_1\cdot v_1^2)(v'\cdot C_t^T) + \sum_{2\leq i\leq d}(e_i\cdot v_1^2)(e_i\cdot C_t^T).$$

Thus, since for any vector $v \in \mathrm{span}(v_2,\dots,v_d)$, $v\cdot C_t^T$ converges to zero, we deduce that $v_1^2\cdot C_t^T$ has the same asymptotic behavior as

$$\frac{e_1\cdot v_1^2}{e_1\cdot v_1}(v_1\cdot C_t^T).$$

Therefore, letting T go to infinity in (5), we can expect $v_1 \cdot C_t^T$ to be the solution of the following stochastic differential equation

$$X_t = \frac{\mu}{m}\int_0^t e^{-\frac{\lambda}{m}(t-s)}ds(v_1\cdot\mathbf{1}) + \frac{1}{m}\sqrt{\frac{e_1\cdot v_1^2}{e_1\cdot v_1}}\int_0^t e^{-\frac{\lambda}{m}(t-s)}\sqrt{X_s}dB_s.$$

This exactly corresponds to a Cox-Ingersoll-Ross process since it can be rewritten

$$dX_t = \frac{\lambda}{m}\big(\frac{\mu}{\lambda}(v_1\cdot\mathbf{1}) - X_t\big)dt + \frac{1}{m}\sqrt{\frac{e_1\cdot v_1^2}{e_1\cdot v_1}}\sqrt{X_t}dB_t, \quad X_0 = 0.$$

Hence, using the decomposition of C_t^T in the basis $(e_1,\ldots,e_d)$ given by

$$C_t^T = \frac{1}{e_1\cdot v_1}(v_1\cdot C_t^T)e_1 + (v'\cdot C_t^T)e_1 + \sum_{2\le i\le d}(e_i\cdot C_t^T)e_i,$$

we finally obtain the following theorem

Theorem 2. *Under the setting and notations of Section 2.3.1 together with Assumptions 2 and 3, the multidimensional process*

$$(C_t^T, B_t^T) = \big(\frac{1}{T}\lambda_{tT}^T, \int_0^{tT}\frac{v_1\cdot dM_s^T}{\sqrt{Tv_1^2\cdot\lambda_s^T}}\big), \quad t\in[0,1]$$

converges in law for the Skorokhod topology to $(\frac{1}{e_1\cdot v_1}Xe_1, B)$ *where* B *is a Brownian motion and* X *satisfies the following (one-dimensional) Cox-Ingersoll-Ross dynamic*

$$dX_t = \frac{\lambda}{m}\big(\frac{\mu}{\lambda}(v_1\cdot\mathbf{1}) - X_t\big)dt + \frac{1}{m}\sqrt{\frac{e_1\cdot v_1^2}{e_1\cdot v_1}}\sqrt{X_t}dB_t, \quad X_0 = 0.$$

Theorem 2, whose rigorous proof is given in Section 4.1, is a general result about the asymptotic behavior of multidimensional nearly unstable Hawkes processes. We see in particular that the non-degeneracy concentrates around the first eigenvector. Also, from Theorem 2, we obtain an immediate corollary given below which will enable us to prove Theorem 1.

2.3.3 Application to our microscopic model

Let us consider a bi-dimensional Hawkes processes sequence $N^T = (N^{T,+}, N^{T,-})$ with intensity $\lambda^T = (\lambda^{T,+}, \lambda^{T,-})$ as in Assumption 1. In this case, the Hawkes processes sequence follows the setting of Section 2.3.1 with $d = 2$,

$$\lambda_1 = \varphi_1 + \beta\varphi_2, \quad \lambda_2 = \varphi_1 - \varphi_2,$$

and

$$v_1 = \begin{pmatrix}1\\ \beta\end{pmatrix}, \quad v_2 = \begin{pmatrix}1\\ -1\end{pmatrix}.$$

We therefore have the following corollary of Theorem 2 which will lead us to the long term limit of our microscopic price model.

Corollary 1. *Under Assumptions 1, 2 and 3, the process*

$$(C_t^{T,+}, C_t^{T,-}, B_t^T) = \Big(\frac{1}{T}\lambda_{tT}^{T,+}, \frac{1}{T}\lambda_{tT}^{T,-}, \int_0^{tT} \frac{dM_s^{T,+} + \beta dM_s^{T,-}}{\sqrt{T(\lambda_s^{T,+} + \beta^2 \lambda_s^{T,-})}}\Big), \quad t \in [0,1]$$

converges in law for the Skorokhod topology to $(\frac{1}{\beta+1}X, \frac{1}{\beta+1}X, B)$ *where* B *is a Brownian motion and* X *satisfies the following (one-dimensional) Cox-Ingersoll-Ross dynamic*

$$dX_t = \frac{\lambda}{m}\Big(\frac{\mu}{\lambda}(\beta+1) - X_t\Big)dt + \frac{1}{m}\sqrt{\frac{1+\beta^2}{1+\beta}}\sqrt{X_t}dB_t, \quad X_0 = 0.$$

Here X essentially corresponds to a limiting volatility process. The Brownian motion in the dynamic of X comes from the limit of B^T, the process defined in (6) and driven by $v_1 \cdot dM_s^T$. In our microscopic model, $M^T = (M^{T,+}, M^{T,-})$. As will be clear from the proof of Theorem 1, the Brownian motion driving the price in Theorem 1 arises from the limiting behavior of $M^{T,+} - M^{T,+}$. Hence, the emergence of leverage effect in the limit is due to the non-zero covariation between $v_1 \cdot dM_s^T$ and $M^{T,+} - M^{T,+}$.

3 From high frequency features to rough volatility

3.1 Encoding Property iv

In Section 2, we have built a microscopic Hawkes-based price model compatible with Properties i, ii and iii. Theorem 1 states that it converges in the long run to a classical Heston model. However, Property iv, that is the wide presence of metaorders on the market, which is a crucial feature of high frequency markets, is not encoded in such models. As explained in [JR16b], this can be translated in the Hawkes framework by considering the model defined by Assumption 1 but under the condition that the kernel matrix exhibits a heavy tail, as observed in practice, see [BJM16, HBB13]. Consequently, we need to replace Assumption 2 in order to get a slowly decreasing behavior for the kernel matrix. This also implies a modification of the asymptotic setting in order to retrieve a non-degenerate scaling limit, see [JR16b]. More precisely, in this section, instead of Assumption 2 we consider the following one.

Assumption 4. *There exist* $\alpha \in (1/2, 1)$ *and* $C > 0$ *such that*

$$\alpha x^\alpha \int_x^\infty \lambda_1(s)ds \underset{x\to\infty}{\to} C.$$

Moreover, for some $\lambda^* > 0$ *and* $\mu > 0$,

$$T^\alpha(1 - a_T) \underset{T\to\infty}{\to} \lambda^* > 0, \quad T^{1-\alpha}\mu_T \underset{T\to\infty}{\to} \mu.$$

Of course, the first eigenvalue under Assumption 1 being $\varphi_1 + \beta\varphi_2$, Assumption 4 on λ_1 can also be expressed in term of the asymptotic behavior of φ_1 and φ_2. Note that in practice, estimated values for α are actually close to $1/2$, see [BJM16, HBB13]. We now give the asymptotic behavior of our price model under Assumption 4.

3.2 The rough macroscopic limit of the high frequency model

Let $\lambda = \alpha\lambda^*/\big(C\Gamma(1-\alpha)\big)$. We have the following result for the long term limit of our microscopic model compatible with Properties i, ii, iii and iv.

Theorem 3. *Under Assumptions 1 and 4, as T tends to infinity, the rescaled microscopic price*

$$\sqrt{\frac{1-a_T}{\mu T^\alpha}}P^T_{tT} = \sqrt{\frac{1-a_T}{\mu T^\alpha}}(N^{T,+}_{tT} - N^{T,-}_{tT}), \quad t \in [0,1],$$

converges in the sense of finite dimensional laws to the following rough Heston model

$$P_t = \frac{1}{1-(\|\varphi_1\|_1 - \|\varphi_2\|_1)}\sqrt{\frac{2}{\beta+1}}\int_0^t \sqrt{Y_s}dW_s,$$

with Y the unique solution of

$$Y_t = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda\big((1+\beta)-Y_s\big)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda\sqrt{\frac{1+\beta^2}{\lambda^*\mu(1+\beta)}}\sqrt{Y_s}dB_s,$$

where (W,B) is a correlated bi-dimensional Brownian motion with

$$d\langle W,B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

Furthermore, the process Y_t has Hölder regularity $\alpha - 1/2 - \varepsilon$ for any $\varepsilon > 0$.

Remark 1. *Theorem 3 states the convergence in the sense of finite dimensional laws and not in Skorokhod topology. The latter does not hold in general. Nevertheless, we have the convergence for the Skorokhod topology of the integrated price*

$$\int_0^t \sqrt{\frac{1-a_T}{\mu T^\alpha}}P^T_{sT}ds$$

to $\int_0^t P_s ds$. Such convergence also holds for the rescaled microscopic price itself under the additional assumption $\varphi_1 = \varphi_2$.

Remark 2. *In the heavy-tailed case, the tick size is taken of order $T^{-\alpha}$. The condition $T^{1-\alpha}\mu_T \to \mu > 0$ means that the exogenous part still does not vanish in the limit since we have exogenous moves of size $O(T^{-\alpha})$ for $\mu_T T = O(T^\alpha)$ times on average in a unit interval.*

Compared to Theorem 1, the only significant difference in the limiting dynamic here is the kernel $(t-s)^{\alpha-1}$ appearing in the two integrals in the volatility process Y_t. Such kernel is similar to that which allows to define a fractional Brownian motion. Indeed, recall that a fractional Brownian motion W^H with Hurst parameter $H \in (0,1)$ can be built through the Mandelbrot-van Ness representation

$$W^H_t = \frac{1}{\Gamma(H+1/2)}\int_{-\infty}^0 \big((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\big)dW_s + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-\frac{1}{2}}dW_s. \qquad (7)$$

Thus, the tail exponent α in Theorem 3 corresponds to a Hurst parameter $\alpha - 1/2$. Our α belonging to $(1/2, 1)$ and in practice being close to $1/2$, the Hurst parameter associated to our limiting volatility is (much) smaller than $1/2$. Therefore, the volatility trajectories are much rougher than that of a Brownian motion and this is why we call our process rough Heston model.

Hence, we have finally shown that when put together in a simple but sufficiently realistic framework, Properties i, ii, iii and iv, which are obvious stylized facts of market microstructure, lead to rough volatility and leverage effect. To our knowledge, this is the first result explaining from an agent-based point of view (although in reduced form) the rough stochastic nature of volatility and in addition leverage effect.

The proof of Theorem 3 is given in Section 4.4. As for Theorem 1 it is based on a result on general multidimensional Hawkes processes (but here with heavy tail) which we explain in the next section.

3.3 Convergence of heavy-tailed nearly unstable multidimensional Hawkes processes

We give in this section a general result for the asymptotic behavior of heavy-tailed nearly unstable multidimensional Hawkes processes. This result will be the key to the proof of Theorem 3. We consider the same setting as in Section 2.3.1 but we work here under Assumption 4. This will imply that the result we can obtain here is slightly weaker than that of Theorem 2. In particular the sequence of intensities is typically not tight and thus cannot converge. However, the same kind of non-rigorous computations as in Section 2.3.2 still enables us to obtain intuition about the result as explained below.

3.3.1 Intuition for the result and theorem

As in Section 2.3.2, we consider a suitable renormalization of the intensity, namely we work with the process

$$C_t^T = \frac{1 - a_T}{\mu_T} \lambda_{tT}^T, \quad t \in [0, 1].$$

Remark that in the setting of Section 2.3.2, the intensity is multiplied by $1/T$. This can be done since under Assumption 2 the factor $(1 - a_T)/\mu_T$ is of order $1/T$. This is no longer the case under Assumption 4.

Following the same computations as in Section 2.3.2, we obtain

$$v_i \cdot C_t^T = (1 - a_T)(v_i \cdot \mathbf{1}) + (v_i \cdot \mathbf{1}) \int_0^t \rho_i^T (t - s) ds + \int_0^t \rho_i^T (t - s)(v_i \cdot d\tilde{M}_s^T),$$

where $\rho_i^T = T(1 - a_T)\psi_i^T(T.)$ and $\tilde{M}_t^T = M_{tT}^T / (T\mu_T)$, which is a martingale such that $\mathbb{E}[\langle \tilde{M}^T, \tilde{M}^T \rangle_t]$ is bounded. Hence we need to study the behavior of ρ_i^T.

In the same way as in Section 2.3.2, using its Laplace transform we get that ρ_i^T should vanish as T goes to infinity for $i \geq 2$. For $i = 1$, we have

$$\hat{\rho}_1^T(z) = \int_0^\infty \rho_1^T(x)e^{-zx}dx = (1-a_T)\hat{\psi}_1^T(z/T) = (1-a_T)\frac{a_T\hat{\lambda}_1(z/T)}{1-a_T\hat{\lambda}_1(z/T)}.$$

Then, integrating by parts and using that $\|\lambda_1\| = 1$, we get

$$\hat{\lambda}_1(z) = \int_0^\infty \lambda_1(x)e^{-zx}dx = 1 - z\int_0^\infty \int_x^\infty \lambda_1(u)du e^{-zx}dx.$$

Therefore,

$$\hat{\lambda}_1(z) = 1 - z^\alpha \int_0^\infty (\frac{x}{z})^\alpha \int_{x/z}^\infty \lambda_1(u)du x^{-\alpha}e^{-x}dx.$$

Hence, using Assumption 4 together with the dominated convergence theorem we obtain

$$\hat{\lambda}_1(z) = 1 - \frac{C}{\alpha}\Gamma(1-\alpha)z^\alpha + \underset{z\to 0}{o}(z).$$

From this, we easily deduce that for $z > 0$,

$$\hat{\rho}_1^T(z) \to \frac{\lambda}{\lambda + z^\alpha},$$

which is the Laplace transform of the Mittag-Leffler density function $f^{\alpha,\lambda}$ defined in Appendix I.C. Consequently, using the same arguments as in Section 2.3.2, we get that C_t^T should essentially satisfy

$$C_t^T \underset{T\to\infty}{\to} \frac{1}{e_1 \cdot v_1} Y_t e_1,$$

where Y is solution of the following rough stochastic differential equation

$$Y_t = (v_1 \cdot \mathbf{1})F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu\lambda^*}}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}}\int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s}dB_s,$$

with $F^{\alpha,\lambda}(t) = \displaystyle\int_0^t f^{\alpha,\lambda}(s)ds$. In fact, this last equation is equivalent to that of a rough Cox-Ingersoll-Ross process,

$$Y_t = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(v_1 \cdot \mathbf{1} - Y_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\frac{\lambda}{\sqrt{\mu\lambda^*}}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}}\sqrt{Y_s}dB_s,$$

see Proposition 9.

Thus, the preceding computations seem to indicate that in the heavy tailed case, the renormalized intensity process should converge to a rough Cox-Ingersoll-Ross process. Contrary to the light tailed case, this intuition is actually not correct in general when the kernel matrix has a slowly decreasing behavior. However, it still holds provided we consider the integrated intensity instead of the intensity itself. We now give the rigorous result.

For $t \in [0,1]$, let us define

$$X_t^T = \frac{1-a_T}{T^\alpha \mu} N_{tT}^T, \quad \Lambda_t^T = \frac{1-a_T}{T^\alpha \mu} \int_0^{tT} \lambda_s^T ds, \quad Z_t^T = \sqrt{\frac{T^\alpha \mu}{1-a_T}} (X_t^T - \Lambda_t^T). \tag{8}$$

We have the following theorem.

Theorem 4. *Under the setting and notations of Section 2.3.1 together with Assumption 4, the process* $(\Lambda_t^T, X_t^T, Z_t^T)_{t\in[0,1]}$ *defined by* (8) *converges in law for the Skorokhod topology to* (Λ, X, Z) *where*

$$\Lambda_t = X_t = \frac{1}{e_1 \cdot v_1} (\int_0^t Y_s ds) e_1$$

and for $1 \leq i \leq d$,

$$Z_t^i = \int_0^t \sqrt{\frac{e_{1,i}}{e_1 \cdot v_1} Y_s} dB_s^i,$$

where $(B^1, \dots, B^d)$ *is a* d*-dimensional Brownian motion and* Y *is the unique solution of the following rough stochastic differential equation*

$$Y_t = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda (v_1 \cdot \mathbf{1} - Y_s) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \frac{\lambda}{\sqrt{\mu\lambda^*}} \sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}} \sqrt{Y_s} dB_s,$$

with

$$B = \frac{1}{\sqrt{e_1 \cdot v_1^2}} \sum_{i=1}^d \sqrt{e_{1,i} v_{1,i}^2} B^i,$$

and $\lambda = \alpha\lambda^* / \big(C\Gamma(1-\alpha)\big)$*. Furthermore,* Y *has Hölder regularity* $\alpha - \frac{1}{2} - \varepsilon$ *for any* $\varepsilon > 0$.

The rigorous proof of Theorem 4 is given in Section 4.3.

3.3.2 Application to our microscopic model

As for Theorem 2, Theorem 4 has an immediate corollary which will be crucial in the proof of Theorem 3. Let us consider a bi-dimensional Hawkes processes sequence $N^T = (N^{T,+}, N^{T,-})$ with intensity $\lambda^T = (\lambda^{T,+}, \lambda^{T,-})$ as in Assumption 1. We have the following result.

Corollary 2. *Under Assumptions 1 and 4, the process* $(\Lambda_t^T, X_t^T, Z_t^T)_{t\in[0,1]}$ *defined by* (8) *converges in law for the Skorokhod topology to* (X, X, Z) *where*

$$X_t = \frac{1}{\beta+1} \int_0^t Y_s ds \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad Z_t = \int_0^t \sqrt{\frac{1}{\beta+1} Y_s} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix},$$

where (B^1, B^2) is a bi-dimensional Brownian motion and Y is the unique solution of the following rough stochastic differential equation

$$Y_t = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda((1+\beta)-Y_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda\sqrt{\frac{1+\beta^2}{\lambda^*\mu(1+\beta)}}\sqrt{Y_s}dB_s,$$

with

$$B = \frac{B^1+\beta B^2}{\sqrt{1+\beta^2}}.$$

4 Proofs

From now on, c denotes a positive constant independent of T that may vary from line to line.

4.1 Proof of Theorem 2

In this proof, which is quite inspired by [JR15], the notations defined in Section 2.3.2 are in force. We start with a lemma often used in the sequel.

4.1.1 A useful lemma

We have the following result.

Lemma 1. *Let $f^T : \mathbb{R}_+ \to \mathbb{R}$ be a sequence of measurable functions such that for some $c > 0$ and any $x_1 \geq 0$, $x_2 \in \mathbb{R}$, $x_3 \geq 0$, $x_4 \geq 0$ and $T > 0$,*

a) $f^T \in \mathbb{L}^1(\mathbb{R}_+) \cap \mathbb{L}^2(\mathbb{R}_+)$ *and* $\int_{x\geq 0} |f^T(x)|^2 dx \underset{T\to\infty}{\to} 0$,
b) $|f^T(x_1)| \leq c$,
c) $|\hat{f}^T(x_2)| \leq c(1 \wedge \frac{1}{|x_2|})$,
d) $|f^T(x_3) - f^T(x_4)| \leq cT|x_3 - x_4|$.

Then, under the setting of Section 2.3.1 together with Assumptions 2 and 3, the process

$$\left(\int_0^t f^T(t-s)d\overline{M}_s^T\right)_{t\in[0,1]}$$

converges to zero in probability as T goes to infinity, uniformly over compact sets (u.c.p.).

The proof of Lemma 1 is given in Appendix I.A.1.

4.1.2 Convergence of $v_i \cdot C^T$ for $i \in \{2, \ldots, d\}$

We now consider the convergence of $C^T = \frac{1}{T}\lambda^T_{.T}$ on the vector space $\mathrm{span}(v_2, \ldots, v_d)$. The following proposition holds.

Proposition 1. *Let $2 \leq i \leq d$. Under the setting of Section 2.3.1 together with Assumptions 2 and 3, $v_i \cdot C^T$ converges u.c.p. to zero as T goes to infinity.*

Proof:

Recall first Equation (4)

$$v_i \cdot C_t^T = \frac{\mu}{T}(v_i \cdot \mathbf{1}) + \mu(v_i \cdot \mathbf{1}) \int_0^t \psi_i^T(T(t-s))ds + \int_0^t \psi_i^T(T(t-s))(v_i \cdot d\overline{M}_s^T).$$

To get the result, it is therefore enough to show that the family of functions $(\psi_i^T(T.))_{T>0}$ satisfies the four points of Lemma 1. Point *b*) is easily obtained from the fact that $v_i^\top.\psi^T = \psi_i^T v_i^\top$ together with the uniform boundedness of ψ^T due to Assumption 3.

Now remark that from Assumption 3, we deduce that $\lambda_i(x)$ tends to zero as x goes to infinity. Then, using integration by parts on the Fourier transform of λ_i together with Assumption 3, we obtain

$$|\hat{\lambda}_i(\omega)| \leq \big((|\lambda_i(0)| + \int_0^\infty |\lambda_i'(x)|dx)\frac{1}{|\omega|}\big) \wedge \|\lambda_i\|_1. \tag{9}$$

Point *c*) follows using that

$$|\hat{\psi}_i^T(T.)(\omega)| = \frac{|a_T \hat{\lambda}_i(\omega/T)|}{|T\big(1 - a_T\hat{\lambda}_i(\omega/T)\big)|} \leq \frac{|a_T\hat{\lambda}_i(\omega/T)|}{T(1-\|\lambda_i\|_1)} \leq c(1 \wedge \frac{1}{|\omega|}).$$

We also obtain from the previous inequality that $\hat{\psi}_i^T(T.)$ is square-integrable and so is $\psi_i^T(T.)$. Moreover by Parseval equality, we have

$$\int_{x\geq 0} |\psi_i^T(Tx)|^2 dx = \frac{1}{2\pi}\int_{\omega\in\mathbb{R}} |\hat{\psi}_i^T(T.)(\omega)|^2 d\omega \leq c\int_{\omega\in\mathbb{R}} \frac{|\hat{\lambda}_i(\omega/T)|^2}{T^2(1-\|\lambda_i\|_1)^2} d\omega \leq \frac{c}{T}\int_{z\in\mathbb{R}} |\hat{\lambda}_i(z)|^2 dz.$$

Since $\hat{\lambda}_i$ is square-integrable, the right hand side of the last inequality tends to zero and thus *a*) is obtained.

Finally *d*) is shown using that $\psi_i^T = a_T\lambda_i + a_T\lambda_i * \psi_i^T$ to write

$$\begin{aligned}|(\psi_i^T)'(Tx)| &= T|a_T\lambda_i'(Tx) + a_T(\lambda_i' * \psi_i^T)(Tx) + a_T\lambda_i(0)\psi_i^T(Tx)| \\ &\leq T(\|\lambda_i'\|_\infty + \|\lambda_i'\|_1\|\psi_i^T\|_\infty + |\lambda_i(0)|\|\psi_i^T\|_\infty).\end{aligned}$$

□

4.1.3 Convergence of $v_1 \cdot C^T$

We have just shown that $v_i \cdot C^T$ tends to zero for $i \in \{2,\ldots,d\}$. The fact that $\|\lambda_i\|_1 < 1$ for $i \in \{2,\ldots,d\}$ was crucial in order to obtain this result. We now treat the term $v_1 \cdot C^T$, recalling that $\|\lambda_1\|_1 = 1$. We have the following result.

Proposition 2. *Under the setting of Section 2.3.1 together with Assumptions 2 and 3, the process* $(v_1 \cdot C_t^T, B_t^T)_{t\in[0,1]}$ *converges in law for the Skorokhod topology to* (X, B) *where* B *is a Brownian motion and* X *satisfies the following Cox-Ingersoll-Ross dynamic*

$$dX_t = \frac{\lambda}{m}(\frac{\mu}{\lambda}(v_1 \cdot \mathbf{1}) - X_t)dt + \frac{1}{m}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}}\sqrt{X_t}dB_t, \quad X_0 = 0.$$

Proof:

1. Rewriting $v_1 \cdot C^T$ Let

$$S_t^T = \sum_{i=2}^{d}(e_i \cdot C_t^T)(e_i \cdot v_1^2) + (v' \cdot C_t^T)(e_1 \cdot v_1^2).$$

From Proposition 1, we get that S_t^T tends u.c.p. to zero. We have

$$v_1^2 \cdot C_t^T = S_t^T + \frac{e_1 \cdot v_1^2}{e_1 \cdot v_1} v_1 \cdot C_t^T,$$

which together with (5) leads to the following expression for $v_1 \cdot C^T$

$$v_1 \cdot C_t^T = \frac{\mu}{T}(v_1 \cdot \mathbf{1}) + \mu(v_1 \cdot \mathbf{1})\int_0^t \psi_1^T(Ts)ds + \int_0^t \psi_1^T\big(T(t-s)\big)\sqrt{S_s^T + \frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}(v_1 \cdot C_s^T)}dB_s^T.$$

2. Convergence of $\psi_1^T(T.)$ For $x \geq 0$, let us define

$$f^T(x) = \psi_1^T(Tx) - \frac{1}{m}\exp(-\frac{\lambda x}{m}).$$

We have seen in Section 2.3.2 that $f^T(x)$ should be close to zero as T goes to infinity. More precisely, we have the following proposition whose proof is given in Corollaries 4.1, 4.2, 4.3 and 4.4 in [JR15]

Proposition 3. *Under the setting of Section 2.3.1 together with Assumptions 2 and 3,* f^T *satisfies Properties a), b), c) and d) of Lemma 1.*

3. The Cox-Ingersoll-Ross like dynamic of $v_1 \cdot C^T$ We can write

$$v_1 \cdot C_t^T = R_t^T + \frac{\mu}{m}(v_1 \cdot \mathbf{1})\int_0^t \exp(-\frac{\lambda s}{m})ds + \frac{1}{m}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}}\int_0^t \exp(-\frac{\lambda(t-s)}{m})\sqrt{v_1 \cdot C_s^T}dB_s^T,$$

with

$$\begin{aligned} R_t^T = {} & \frac{\mu}{T}(v_1 \cdot \mathbf{1}) + \mu(v_1 \cdot \mathbf{1})\int_0^t f^T(s)ds + \int_0^t f^T(t-s)(v_1 \cdot d\overline{M}_s^T) \\ & + \frac{1}{m}\int_0^t \exp(-\frac{\lambda(t-s)}{m})\Big(\sqrt{S_s^T + \frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}(v_1 \cdot C_s^T)} - \sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}(v_1 \cdot C_s^T)}\Big)dB_s^T. \end{aligned} \tag{10}$$

Then, using integration by parts, we get that

$$\int_0^t \exp(-\frac{\lambda(t-s)}{m})\sqrt{v_1 \cdot C_s^T}\,dB_s^T$$

is equal to

$$\int_0^t \sqrt{v_1 \cdot C_s^T}\,dB_s^T - \frac{\lambda}{m}\int_0^t\int_0^s \exp(-\frac{\lambda(s-u)}{m})\sqrt{v_1 \cdot C_u^T}\,dB_u^T\,ds.$$

This can be rewritten

$$\int_0^t \sqrt{v_1 \cdot C_s^T}\,dB_s^T - \lambda\sqrt{\frac{e_1 \cdot v_1}{e_1 \cdot v_1^2}}\int_0^t v_1 \cdot C_s^T - R_s^T - \frac{\mu}{\lambda}(v_1 \cdot \mathbf{1})\big(1-\exp(-\frac{\lambda s}{m})\big)ds.$$

Consequently,

$$v_1 \cdot C_t^T = U_t^T + \int_0^t \frac{\lambda}{m}\big(\frac{\mu}{\lambda}(v_1 \cdot \mathbf{1}) - v_1 \cdot C_s^T\big)ds + \frac{1}{m}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}}\int_0^t \sqrt{v_1 \cdot C_s^T}\,dB_s^T, \tag{11}$$

with

$$U_t^T = R_t^T + \frac{\lambda}{m}\int_0^t R_s^T\,ds.$$

4. Convergence of U^T We now show that U^T converges u.c.p. to zero. This vanishing behavior comes from that of f^T and S^T. Of course it is enough to prove that R^T converges u.c.p to zero. From Proposition 3 together with Lemma 1, it is obvious that the first three terms in (10) tend to zero. We now treat the last term.

First, remark that

$$|\sqrt{S_s^T + \beta v_1 \cdot C_s^T} - \sqrt{\beta v_1 \cdot C_s^T}| \le \sqrt{|S_s^T|},$$

which tends to zero as T goes to infinity thanks to Proposition 1. Furthermore, observe that since $\langle M^T, M^T\rangle = \mathrm{diag}(\int_0^{\cdot} \lambda^T)$ and $\lambda^T \ge \mu\mathbf{1}$, we have

$$\mathbb{E}\Big[\int_0^{tT} \frac{v_1^2 \cdot dM_s^T}{T v_1^2 \cdot \lambda_s^T}\Big]^2 = \mathbb{E}\Big[\int_0^{tT} \frac{v_1^4 \cdot \lambda_s^T ds}{T^2 (v_1^2 \cdot \lambda_s^T)^2}\Big] \le \mathbb{E}\Big[\int_0^{tT} \frac{v_1^4 \cdot \lambda_s^T ds}{T^2 \mu (v_1^4 \cdot \lambda_s^T)}\Big]. \tag{12}$$

We get that this is smaller than c/T and consequently goes to zero. Therefore, B^T is a sequence of martingales with bounded jumps whose quadratic variation given by [3]

$$[B^T, B^T]_t = t + \int_0^{tT} \frac{v_1^2 \cdot dM_s^T}{T v_1^2 \cdot \lambda_s^T},$$

tends in probability to identity for any fixed t. Using Theorem VIII-3.11 in [JS13], this implies that B^T converges in law towards a Brownian motion B for the Skorokhod topology. Moreover

[3]We use the fact that $[M^T, M^T]_t = \mathrm{diag}(N_t^T)$ and $N_t^T = M_t^T + \int_0^t \lambda_s^T ds$.

by Proposition 3.2 in [JMP89], the sequence B^T satisfies the U.T condition. From Theorem 2.6 in [JMP89], the convergence for the Skorokhod topology to zero of

$$\frac{1}{m}\int_0^t \exp(-\frac{\lambda(t-s)}{m})\big(\sqrt{S_s^T+\frac{e_1\cdot v_1}{e_1\cdot v_1^2}(v_1\cdot C_s^T)}-\sqrt{\frac{e_1\cdot v_1}{e_1\cdot v_1^2}(v_1\cdot C_s^T)}\big)dB_s^T$$

follows and finally we get that U^T tends to zero u.c.p.

5. End of the proof of Proposition 2 We have that $v_1\cdot C_t^T$ can be written as in (11) and furthermore (B^T,U^T) converges in law for the Skorokhod topology to $(B,0)$. Proposition 2 readily follows from Theorem 5.4 in [KP91]. □

4.1.4 End of proof of Theorem 2

Decomposing C^T in the basis $(e_1,\dots,e_d)$,

$$C_t^T=\sum_{i=2}^d (e_i\cdot C_t^T)e_i+(v'\cdot C_t^T)e_1+\frac{1}{e_1\cdot v_1}(v_1\cdot C_t^T)e_1,$$

we immediately obtain Theorem 2 from Proposition 1 together with Proposition 2.

4.2 Proof of Theorem 1

4.2.1 Convenient rewriting of P^T

We start by writing conveniently our rescaled price P_{tT}^T/T. We have

$$\frac{1}{T}P_{tT}^T=\frac{N_{tT}^{T,+}-N_{tT}^{T,-}}{T}=\int_0^{tT}\frac{dM_s^{T,+}-dM_s^{T,-}}{\sqrt{T(\lambda_s^{T,+}+\lambda_s^{T,-})}}\sqrt{\frac{\lambda_s^{T,+}+\lambda_s^{T,-}}{T}}+\int_0^{tT}\frac{\lambda_s^{T,+}-\lambda_s^{T,-}}{T}ds.$$

Furthermore,

$$\begin{aligned}\lambda_t^{T,+}-\lambda_t^{T,-}&=\int_0^t a_T(\varphi_1(t-s)-\varphi_2(t-s))(dN_s^{T,+}-dN_s^{T,-})\\&=\int_0^t a_T\lambda_2(t-s)(dM_s^{T,+}-dM_s^{T,-})+\int_0^t a_T\lambda_2(t-s)(\lambda_s^{T,+}-\lambda_s^{T,-})ds.\end{aligned}$$

Thus, from Lemma 2, we obtain

$$\lambda_t^{T,+}-\lambda_t^{T,-}=\int_0^t \psi_2^T(t-s)(dM_s^{T,+}-dM_s^{T,-}).$$

Then, using Fubini theorem, we get

$$\int_0^x \lambda_s^{T,+}-\lambda_s^{T,-}ds=\int_0^x\big(\int_0^{x-s}\psi_2^T(u)du\big)(dM_s^{T,+}-dM_s^{T,-}).$$

Hence our rescaled price process P^T_{tT}/T can finally be written

$$\int_0^t \sqrt{C_s^{T,+}+C_s^{T,-}}dW_s^T - \int_0^t\int_{T(t-s)}^{\infty}\psi_2^T(u)du(d\overline{M}_s^{T,+}-d\overline{M}_s^{T,-})+\int_0^{\infty}\psi_2^T(u)du(\overline{M}_t^{T,+}-\overline{M}_t^{T,-}),$$

with

$$W_t^T=\int_0^{tT}\frac{dM_s^{T,+}-dM_s^{T,-}}{\sqrt{T(\lambda_s^{T,+}+\lambda_s^{T,-})}}.$$

Since $\int_0^\infty \psi_2^T(u)du = \frac{1}{1-a_T\int_0^\infty \lambda_2(s)ds}$, we obtain

$$\frac{1}{T}P^T_{tT}=\frac{1}{1-a_T(\|\varphi_1\|_1-\|\varphi_2\|_2)}\int_0^t\sqrt{C_s^{T,+}+C_s^{T,-}}dW_s^T-R_t^T, \tag{13}$$

with

$$R_t^T=\int_0^t\int_{T(t-s)}^{\infty}\psi_2^T(u)du(d\overline{M}_s^{T,+}-d\overline{M}_s^{T,-}).$$

4.2.2 Convergence of R^T

We have the following proposition.

Proposition 4. *Under Assumptions 1, 2 and 3, R^T tends u.c.p. to zero.*

Proof.

From Lemma 1, it is enough to show that the sequence of functions

$$g^T(x)=\int_{Tx}^{\infty}\psi_2^T(u)du$$

satisfies Properties *a*), *b*), *c*) and *d*) of Lemma 1. The fact that *b*) holds is obvious since

$$|g^T(z)|\leq\int_0^{+\infty}|\psi_2^T(x)|dx\leq\frac{||\lambda_2||_1}{1-||\lambda_2||_1}.$$

Then we remark that

$$\hat{g}^T(z)=\int_{a\geq 0}\psi_2^T(a)\frac{e^{iza/T}-1}{iz}da,$$

which shows that Property *c*) holds. Property *d*) is obtained from the fact that

$$|(g^T)'(x)|=T|\psi_2^T(Tx)|\leq cT.$$

Finally, we use Fubini theorem to write

$$\int_{x\geq 0}|g^T(x)|^2dx=\int_{x\geq 0;a,b>Tx}\psi_2^T(a)\psi_2^T(b)dadbdx=\frac{1}{T}\int_{a,b\geq 0}(a\wedge b)\psi_2^T(a)\psi_2^T(b)dadb.$$

Consequently,

$$\int_{x\geq 0} |g^T(x)|^2 dx \leq \frac{1}{T}\int_{a\geq 0} a|\psi_2^T(a)|da \int_{b\geq 0} |\psi_2^T(b)|db \leq \frac{c}{T}\sum_{k\geq 1}\int_{a\geq 0} a|\lambda_2|^{*k}(a)da.$$

By recursion, we get that for $k \geq 1$,

$$\int_{a\geq 0} a|\lambda_2|^{*k}(a)da = k||\lambda_2||_1^{k-1}\int_{a\geq 0} a|\lambda_2|(a)da < \infty.$$

Eventually

$$\int_{x\geq 0} |g^T(x)|^2 dx \leq c/T$$

and a) easily follows. □

4.2.3 Convergence of (W^T, B^T)

In the same way as for the quadratic variation of B^T in the proof of Theorem 2, we easily get the following convergence in probability for any fixed $t \in [0,1]$

$$[W^T, W^T]_t \underset{T\to\infty}{\to} t, \quad [B^T, B^T]_t \underset{T\to\infty}{\to} t.$$

Moreover, we have the following proposition.

Proposition 5. *Under Assumptions 1, 2 and 3,*

$$[W^T, B^T]_t \underset{T\to\infty}{\to} \frac{1-\beta}{\sqrt{2(1+\beta^2)}}t$$

in probability, for any fixed $t \in [0,1]$.

Proof.

Using $[M^T, M^T] = \mathrm{diag}(N^T)$, we write

$$\begin{aligned}
[W^T, B^T]_t &= \int_0^{tT} \frac{dN_s^{T,+} - \beta dN_s^{T,-}}{T\sqrt{\lambda_s^{T,+} + \lambda_s^{T,-}}\sqrt{\lambda_s^{T,+} + \beta^2\lambda_s^{T,-}}} \\
&= \int_0^{t} \frac{C_s^{T,+} - \beta C_s^{T,-}}{\sqrt{C_s^{T,+} + C_s^{T,-}}\sqrt{C_s^{T,+} + \beta^2 C_s^{T,-}}} ds + \varepsilon_t^T,
\end{aligned}$$

with

$$\varepsilon_t^T = \int_0^{tT} \frac{dM_s^{T,+} - \beta dM_s^{T,-}}{T\sqrt{\lambda_s^{T,+} + \lambda_s^{T,-}}\sqrt{\lambda_s^{T,+} + \beta^2\lambda_s^{T,-}}}.$$

Since $\langle M^T, M^T\rangle = \mathrm{diag}(\int_0^{\cdot} \lambda^T)$ and $\lambda^T \geq \mu\mathbf{1}$, we easily get

$$\mathbb{E}[(\varepsilon_t^T)^2] = \mathbb{E}[\int_0^{tT} \frac{1}{T^2(\lambda_s^{T,+} + \lambda_s^{T,-})}] \leq \frac{1}{2\mu T} \underset{T\to\infty}{\to} 0.$$

Furthermore, from Corollary 1, $(C^{T,+}, C^{T,-})$ converges in law for the Skorokhod topology to $\left(\frac{1}{\beta+1}X, \frac{1}{\beta+1}X\right)$. Using Skorokhod's representation theorem, there exists a probability space on which one can duplicate in law the sequence (Z^T, X^T) so that it converges almost surely for the Skorohod topology to a random variable with the same law as (Z, X). The set of zeros of a Cox-Ingersoll-Ross process on a finite interval being of Lebesgue measure zero, we deduce that almost surely,

$$\frac{C_t^{T,+} - \beta C_t^{T,-}}{\sqrt{C_t^{T,+} + C_t^{T,-}}\sqrt{C_t^{T,+} + \beta^2 C_t^{T,-}}}$$

tends for almost every $t \in [0,1]$ to

$$\frac{1-\beta}{\sqrt{2(1+\beta^2)}}.$$

Thus we deduce by dominated convergence theorem[4] that almost surely, for all $t \in [0,1]$,

$$\int_0^t \frac{C_s^{T,+} - \beta C_s^{T,-}}{\sqrt{C_s^{T,+} + C_s^{T,-}}\sqrt{C_s^{T,+} + \beta^2 C_s^{T,-}}} ds \underset{T\to\infty}{\to} \frac{1-\beta}{\sqrt{2(1+\beta^2)}} t.$$

Therefore, for any $t \in [0,1]$, we have the following convergence in probability

$$[B^T, W^T]_t \underset{T\to\infty}{\to} \frac{1-\beta}{\sqrt{2(1+\beta^2)}} t.$$

□

4.2.4 End of the proof of Theorem 1

Consider (13). From Proposition 4, R^T tends to zero. Then using Theorem VIII-3.11 in [JS13] together with Proposition 5, we obtain that (W^T, B^T) converges in law for the Skorokhod topology to a correlated bi-dimensional Brownian motion (W, B) such that

$$\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}} t.$$

Furthermore, from Corollary 1 we get that $(\sqrt{C^{T,+} + C^{T,-}}, B^T)$ converges in law for the Skorokhod topology to $(\sqrt{\frac{2}{\beta+1}X}, B)$, where X is a Cox-Ingersoll-Ross process driven by B and defined in Corollary 1. Moreover, W^T being a martingale with bounded jumps, it satisifies the U.T condition from Proposition 3.2 in [JMP89]. Using Theorem 2.6 in [JMP89], we deduce that

$$\int_0^t \sqrt{C_s^{T,+} + C_s^{T,-}} dW_s^T$$

converges in law for the Skorokhod topology to

$$\int_0^t \sqrt{\frac{2X_s}{1+\beta}} dW_s,$$

which ends the proof.

[4] Notice that $\forall (x,y) \in (\mathbb{R}_+^2)^*, \quad |\frac{x-\beta y}{\sqrt{x+y}\sqrt{x+\beta^2 y}}| \leq 1$ by Cauchy-Schwarz inequality.

4.3 Proof of Theorem 4

We now give the proof of our theorem on the convergence of general nearly unstable Hawkes processes with heavy tail. This proof is quite inspired from [JR16b].

4.3.1 C-tightness of (Λ^T, X^T, Z^T)

We have the following proposition.

Proposition 6. *Under the setting of Section 2.3.1 together with Assumption 4, the sequence* (Λ^T, X^T, Z^T) *is C-tight and*

$$\sup_{t\in[0,1]} \|\Lambda_t^T - X_t^T\| \underset{T\to\infty}{\to} 0$$

in probability. Moreover if (X, Z) *is a possible limit point of* (X^T, Z^T)*, then* Z *is a continuous martingale with* $[Z, Z] = \mathrm{diag}(X)$.

Proof:

1. C-tightness of X^T and Λ^T Recall that as in (3), we can write

$$\lambda_t^T = \mu_T \mathbf{1} + \mu_T \int_0^t \psi^T(t-s)ds.\mathbf{1} + \int_0^t \psi^T(t-s).dM_s^T.$$

Using that $\int_0^{\cdot} (f * g) = (\int_0^{\cdot} f) * g$, we get

$$\mathbb{E}[N_T^T] = \mathbb{E}[\int_0^T \lambda_s^T ds] = T\mu_T \mathbf{1} + \mu_T \int_0^T s\psi^T(T-s)ds.\mathbf{1}.$$

Consequently,

$$\mathbf{1}\cdot\mathbb{E}[N_T^T] = T\mu_T d + \mu_T \mathbf{1}\cdot\big((\int_0^T s\psi^T(T-s)ds).\mathbf{1}\big)$$

and therefore

$$\mathbf{1}\cdot\mathbb{E}[N_T^T] \le cT\mu_T\big(1 + \mathscr{S}(\int_0^\infty \psi^T(s)ds)\big) \le c\frac{T\mu_T}{1-a_T} \le cT^{2\alpha}.$$

Thus, we obtain that

$$\mathbb{E}[X_1^T] = \mathbb{E}(\Lambda_1^T) \le c.$$

Each component of X^T and Λ^T being increasing, we deduce the tightness of each component of (X^T, Λ^T). Furthermore, the maximum jump size of X^T and Λ^T being $\frac{1-a_T}{T^\alpha \mu}$ which goes to zero, the C-tightness of (X^T, Λ^T) is obtained from Prop.VI-3.26 in [JS13].

2. C-tightness of Z^T It is easy to check that

$$\langle Z^T, Z^T\rangle = \mathrm{diag}(\Lambda^T),$$

which is C-tight. From Theorem VI-4.13 in [JS13], this gives the tightness of Z^T. The maximum jump size of Z^T vanishing as T goes to infinity, we obtain that Z^T is C-tight.

3. Convergence of $X^T - \Lambda^T$ We have

$$X_t^T - \Lambda_t^T = \frac{1-a_T}{T^\alpha \mu} M_{tT}^T.$$

From Doob's inequality, we get for each component

$$\mathbb{E}[\sup_{t\in[0,1]} |\Lambda_t^T - X_t^T|^2] \leq cT^{-4\alpha}\mathbb{E}[M_T^T]^2.$$

Since $[M^T, M^T] = \text{diag}(N^T)$, we deduce

$$\mathbb{E}[\sup_{t\in[0,1]} |\Lambda_t^T - X_t^T|^2] \leq cT^{-2\alpha}.$$

This gives the uniform convergence to zero in probability of $X^T - \Lambda^T$.

4. Limit of Z^T Let (X,Z) be a possible limit point of (X^T, Z^T). We know that (X,Z) is continuous and from Corollary IX-1.19 of [JS13], Z is a local martingale. Moreover, since

$$[Z^T, Z^T] = \text{diag}(X^T),$$

using Theorem VI-6.26 in [JS13], we get that $[Z,Z]$ is the limit of $[Z^T, Z^T]$ and $[Z,Z] = \text{diag}(X)$. By Skorokhod's representation theorem and Fatou's lemma, the expectation of $[Z,Z]$ is finite and therefore Z is a martingale. □

4.3.2 Convergence of $v_i \cdot X^T$ for $i \geq 2$

Here also, we observe a vanishing behavior in the direction of the eigenvectors v_i for $i \geq 2$. More precisely, we have the following result.

Proposition 7. *Under the setting of Section 2.3.1 together with Assumption 4, if X is a possible limit point of X^T, then for $i \geq 2$ we have $v_i \cdot X = 0$.*

Proof:
From (3), using Fubini theorem together with the fact that $\int_0^{\cdot}(f * g) = (\int_0^{\cdot} f) * g$, we get

$$\int_0^t \lambda_s^T ds = t\mu_T \mathbf{1} + \mu_T \int_0^t s\psi^T(t-s)ds.\mathbf{1} + \int_0^t \psi^T(t-s).M_s^T ds.$$

Therefore, for $t \in [0,1]$, we have the decomposition

$$\Lambda_t^T = T_1 + T_2 + T_3, \tag{14}$$

with

$$T_1 = (1-a_T)tu_T\mathbf{1},$$

$$T_2 = T(1-a_T)u_T\int_0^t s\psi^T\big(T(t-s)\big)ds.\mathbf{1},$$

$$T_3 = T^{1-\alpha/2}\sqrt{\frac{1-a_T}{\mu}}\int_0^t \psi^T\big(T(t-s)\big).Z_s^T\,ds,$$

with $u_T = \mu_T/(\mu T^{\alpha-1})$ tending to one.

Now recall that for $1 \le i \le d$,

$$\psi_i^T = \sum_{k\ge1} a_T^k \lambda_i^{*k}, \quad \rho_i^T = T(1-a_T)\psi_i^T(T.),$$

and define

$$F_i^T = \int_0^{\cdot} \rho_i^T(s)ds.$$

For $i \ge 2$, using that

$$|F_i^T(t)| \le \int_0^t T(1-a_T)|\psi_i^T(Ts)|ds \le (1-a_T)\int_0^\infty |\psi_i^T(s)|ds \le (1-a_T)\frac{\|\lambda_i\|_1}{1-\|\lambda_i\|_1}.$$

we get the uniform convergence to zero of F_i^T. Thanks to this together with integration by parts, we deduce the convergence to zero of $v_i\cdot T_2$ since

$$v_i\cdot T_2 = u_T(v_i\cdot\mathbf{1})\int_0^t F_i^T(s)ds.$$

For $v_i\cdot T_3$ we write

$$v_i\cdot T_3 = \frac{1}{\sqrt{\mu(1-a_T)T^\alpha}}\int_0^t F_i^T(t-s)(v_i\cdot dZ_s^T).$$

The quadratic variation of Z^T being Λ^T which is uniformly bounded in expectation, we have

$$\mathbb{E}[(v_i\cdot T_3)^2] \le \frac{c}{\mu(1-a_T)T^\alpha}\int_0^t \big(F_i^T(s)\big)^2 ds.$$

The convergence of $v_i\cdot T_3$ to zero follows. Finally, from Proposition 6 we have that if X is a limit point of X^T, then X is also a limit point of Λ^T. Therefore, we obtain $v_i\cdot X = 0$.

□

4.3.3 Convergence of $v_1\cdot X^T$

The term $v_1\cdot X^T$ is the non-vanishing one. Indeed, for (Z,X) a possible limit point of (Z^T,X^T), using the same approach as in [JR16b], we obtain

$$T_2\cdot v_1 \underset{T\to\infty}{\to} (v_1\cdot\mathbf{1})\int_0^t s f^{\alpha,\lambda}(t-s)ds$$

and

$$T_3\cdot v_1 \underset{T\to\infty}{\to} \frac{1}{\sqrt{\lambda^*\mu}}\int_0^t f^{\alpha,\lambda}(t-s)(v_1\cdot Z_s)ds.$$

Then, letting T go to infinity in the decomposition (14) we easily deduce the following proposition.

Proposition 8. *Under the setting of Section 2.3.1 together with Assumption 4, if (Z, X) is a possible limit point of (Z^T, X^T), then the process $v_1 \cdot X$ satisfies the following equation*

$$v_1 \cdot X_t = (v_1 \cdot \mathbf{1}) \int_0^t s f^{\alpha,\lambda}(t-s) ds + \frac{1}{\sqrt{\lambda^* \mu}} \int_0^t f^{\alpha,\lambda}(t-s)(v_1 \cdot Z_s) ds.$$

4.3.4 A first version of Theorem 4

We now prove a version of Theorem 4 where Y is specified in a different way. Let (X, Z) be a possible limit point of (X^T, Z^T). From Proposition 8, in the same way as the proof of Theorem 3.2 in [JR16b], we can show that

$$v_1 \cdot X_t = \int_0^t Y_s ds,$$

where Y satisfies

$$Y_t = (v_1 \cdot \mathbf{1}) F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\lambda^* \mu}} \int_0^t f^{\alpha,\lambda}(t-s)(v_1 \cdot dZ_s).$$

Using Proposition 7 together with the decomposition of X_t in the orthonormal basis $(e_1, \dots, e_d)$,

$$X_t = \sum_{i=2}^{d} (e_i \cdot X_t) e_i + (v' \cdot X_t) e_1 + \frac{1}{e_1 \cdot v_1} (v_1 \cdot X_t) e_1,$$

we get

$$X_t = \frac{1}{e_1 \cdot v_1} (v_1 \cdot X_t) e_1 = \frac{1}{e_1 \cdot v_1} (\int_0^t Y_s ds) e_1.$$

From Proposition 6, we have that

$$[Z, Z] = \mathrm{diag}(X) = \frac{1}{e_1 \cdot v_1} (\int_0^t Y_s ds) \mathrm{diag}(e_1).$$

Thus we can use Theorem V-3.9 in [RY13] to show the existence of a d-dimensional Brownian motion $(B^1, \dots, B^d)$ such that for $1 \le i \le d$,

$$Z_t^i = \frac{1}{\sqrt{e_1 \cdot v_1}} \sqrt{e_{1,i}} \int_0^t \sqrt{Y_s} dB_s^i.$$

Finally, in the same way as the proof of Theorem 3.2 in [JR16b], we obtain that Y satisfies

$$Y_t = (v_1 \cdot \mathbf{1}) F^{\alpha,\lambda}(t) + \sqrt{\frac{e_1 \cdot v_1^2}{\lambda^* \mu (e_1 \cdot v_1)}} \int_0^t f^{\alpha,\lambda}(t-s) \sqrt{Y_s} dB_s, \tag{15}$$

where B is a Brownian motion defined by

$$B = \frac{1}{\sqrt{e_1 \cdot v_1^2}} \sum_{1 \le i \le d} \sqrt{e_{1,i}} v_{1,i} B^i.$$

and that Y has Hölder regularity $\alpha - 1/2 - \varepsilon$, for any $\varepsilon > 0$.

4.3.5 End of the proof of Theorem 4

We eventually provide here the proposition showing that from (15), Y can be written under the form of the rough stochastic differential equation given in Theorem 4 and stating the uniqueness of the solution of this equation. Theorem 4 follows immediately.

Proposition 9. *Let λ, ν, θ be positive constants, $\alpha \in (1/2,1)$ and B a Brownian motion. The process V is solution of the following rough stochastic differential equation*

$$V_t = \theta F^{\alpha,\lambda}(t) + \nu \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s}dB_s \tag{16}$$

if and only if it is solution of

$$V_t = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta - V_s)ds + \frac{\lambda\nu}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{V_s}dB_s. \tag{17}$$

Furthermore, both equations admit a unique weak solution.

Proof:

The existence of a solution to (16) has already been proved deriving (15). Let V be a solution to (16) and write

$$K = I^{1-\alpha}V,$$

where $I^{1-\alpha}$ is the fractional integral operator of order $(1-\alpha)$, see Appendix I.B. Using stochastic Fubini theorem, see for example [Ver12], and integration by parts, we get

$$K_t = \theta\int_0^t I^{1-\alpha}f^{\alpha,\lambda}(u)du + \nu\int_0^t I^{1-\alpha}f^{\alpha,\lambda}(t-u)\sqrt{V_u}dB_u.$$

Moreover, since $I^{1-\alpha}f^{\alpha,\lambda}(t) = \lambda\big(1-F^{\alpha,\lambda}(t)\big)$, see Appendix I.C, using stochastic Fubini theorem, we obtain

$$K_t = \lambda\theta\int_0^t \big(1-F^{\alpha,\lambda}(u)\big)du + \nu\lambda\int_0^t \sqrt{V_u}dB_u - \lambda\int_0^t \nu\int_0^s \sqrt{V_u}f^{\alpha,\lambda}(s-u)dB_u ds.$$

Hence,

$$K_t = \lambda\theta\int_0^t \big(1-F^{\alpha,\lambda}(u)\big)du + \nu\lambda\int_0^t \sqrt{V_u}dB_u - \lambda\int_0^t \big(V_s - \theta F^{\alpha,\lambda}(s)\big)ds$$

and finally

$$K_t = \lambda\int_0^t (\theta - V_u)du + \lambda\nu\int_0^t \sqrt{V_u}dB_u.$$

Now recall that we have

$$V_t = D^{1-\alpha}K_t,$$

where the fractional differentiation operator $D^{1-\alpha}$ is defined in Appendix I.B. Thus we get

$$V_t = \frac{1}{\Gamma(\alpha)}\frac{d}{dt}\int_0^t \lambda\int_0^s (s-u)^{\alpha-1}(\theta - V_u)d_u ds + \frac{1}{\Gamma(\alpha)}\frac{d}{dt}\int_0^t \lambda\nu\int_0^s (s-u)^{\alpha-1}\sqrt{V_u}dB_u ds$$

and finally, again from Fubini theorem,

$$V_t = \frac{1}{\Gamma(\alpha)}\lambda \int_0^t (t-u)^{\alpha-1}(\theta - V_u)du + \frac{1}{\Gamma(\alpha)}\lambda\nu \int_0^t (t-u)^{\alpha-1}\sqrt{V_u}dB_u.$$

Therefore V is solution of (17). Using a straightforward generalization of the main result in [MS15], we deduce the uniqueness of such a solution. □

4.4 Proof of Theorem 3

First, remark that in the same way as in Section 4.2, we can write

$$\sqrt{\frac{1-a_T}{T^\alpha \mu}} P_{tT}^T = \frac{1}{1-a_T(\|\varphi_1\|_1 - \|\varphi_2\|_1)}(Z_t^{T,+} - Z_t^{T,-}) - R_t^T,$$

with

$$R_t^T = \int_0^t \Big(\int_{T(t-s)}^\infty \psi_2^T(u)du\Big)(dZ_s^{T,+} - dZ_s^{T,-}).$$

Using Corollary 2, we deduce that

$$\frac{1}{1-a_T(\|\varphi_1\|_1 - \|\varphi_2\|_1)}(Z_t^{T,+} - Z_t^{T,-})$$

converges in law for the Skorokhod topology to the rough Heston dynamic P defined in Theorem 3.

Note that when $\varphi_1 = \varphi_2$, $R^T = 0$. Thus, in this case, we obtain the convergence in law for the Skorokhod topology of the rescaled microscopic price to P. For the general case, we can prove the convergence of R^T to zero in the sense of finite dimensional laws as follows. We have

$$\mathbb{E}[(R_t^T)^2] \leq c\int_0^t \Big(\int_{Ts}^\infty \psi_2^T(u)du\Big)^2 ds.$$

Let $G = \sum_{k\geq 1} |\varphi_1 - \varphi_2|^{*k}$. Note that $|\psi_2^T| \leq G$ and that G is integrable since $\int_0^\infty |\varphi_1 - \varphi_2| < 1$. Hence

$$\mathbb{E}[(R_t^T)^2] \leq c\Big(\int_0^{T^{-1/2}} \Big(\int_{Ts}^\infty G(u)du\Big)^2 ds + \int_{T^{-1/2}}^1 \Big(\int_{Ts}^\infty G(u)du\Big)^2 ds\Big).$$

Then

$$\mathbb{E}[(R_t^T)^2] \leq c\Big(T^{-1/2}\Big(\int_0^\infty G\Big)^2 + \Big(\int_{T^{1/2}}^\infty G\Big)^2\Big),$$

which vanishes as T tends to infinity. The result follows. □

Remark 3. *Note that*

$$\sup_{t\in[0,1]} \Big|\int_0^t \int_{Ts}^\infty \psi_2^T(u)duds\Big| \leq c\Big(T^{-1/2}\int_0^\infty G + \int_{T^{1/2}}^\infty G\Big),$$

which vanishes as T goes to infinity. Then, using Fubini theorem, we get that

$$\int_0^t R_s^T ds = \int_0^t \int_0^{t-s} \Big(\int_{Tu}^{\infty} \psi_2^T\Big) du (dZ_s^{T,+} - dZ_s^{T,-})$$

converges u.c.p. to zero. Thus, as stated in Remark 1, the integrated rescaled microscopic price converges in law for the Skorokhod topology to $\int_0^t P_s ds$.

I.A Technical results

I.A.1 Proof of Lemma 1.

This result has already been proved in [JR16b] in dimension one. We need to generalize it for $d \geq 2$. Inspection of the proof of Proposition 4.1 in [JR16b] shows that the tightness of

$$H_t^T = \int_0^t f^T(t-s) d\overline{M}_s^T$$

with respect to the Skorokhod topology, holds the same way when the dimension is larger than one. So we just need to check the finite dimensional convergence of H^T to zero. Using that $\langle M^T, M^T \rangle = \int_0^\cdot \lambda^T$, we get

$$\mathbb{E}[\|H_t^T\|_2^2] = \frac{1}{T^2}\mathbb{E}\Big[\int_0^{tT} f^T(t-s/T)^2 \sum_{i=1}^d \lambda_{s,i}^T ds\Big] = \frac{1}{T^2}\int_0^{tT} f^T(t-s/T)^2 \sum_{i=1}^d \mathbb{E}[\lambda_{s,i}^T] ds.$$

Using (3) together with the fact that $v_i^\top . \psi^T = \psi_i^T v_i^\top$, we obtain that for any $i \in \{1, \dots, d\}$ and $s \geq 0$,

$$\mathbb{E}[v_i \cdot \lambda_s^T] = \mu(v_i \cdot \mathbf{1})\big(1 + \int_0^s \psi_i^T(u) du\big).$$

Thus

$$|\mathbb{E}[v_i \cdot \lambda_s^T]| \leq \mu |v_i \cdot \mathbf{1}| \big(1 + \sum_{k \geq 1} \int_0^\infty a_T^k |\lambda_i|^{*k}(u) du\big) \leq \mu |v_i \cdot \mathbf{1}| \frac{1}{1 - a_T \|\lambda_i\|_1} \leq cT.$$

Hence for any $i \in \{1, \dots, d\}$, $\mathbb{E}[\lambda_{s,i}^T] \leq cT$. Therefore

$$\mathbb{E}[\|H_t^T\|_2^2] \leq c \int_0^\infty f^T(s)^2 ds \underset{T \to \infty}{\to} 0$$

and so H_t^T tends in probability to zero giving the finite dimensional convergence of the process.

I.A.2 Wiener-Hopf equations

The following result is used extensively in this work to solve Wiener-Hopf type equations, see for example [BDHM13b].

Lemma 2. *Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^d$ and $\phi : \mathbb{R}_+ \to \mathcal{M}^d(\mathbb{R})$ be a matrix-valued function with integrable components such that $\mathcal{S}(\int_0^\infty \phi(s)ds) < 1$. Then there exists a unique locally bounded function f from $\mathbb{R}_+$ to $\mathbb{R}^d$ solution of*

$$f(t) = g(t) + \int_0^t \phi(t-s).f(s)ds, \quad t \geq 0$$

given by

$$f(t) = g(t) + \int_0^t \psi(t-s).g(s)ds, \quad t \geq 0,$$

where $\psi = \sum_{k\geq 1} \phi^{*k}$.

I.B Fractional integrals and derivatives

The fractional integral of order $r \in (0,1]$ of a function f is defined by

$$I^r f(t) = \frac{1}{\Gamma(r)} \int_0^t (t-s)^{r-1} f(s)ds,$$

whenever the integral exists. Its fractional derivative of order $r \in [0,1)$ is given by

$$D^r f(t) = \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s)ds,$$

whenever it exists.

I.C Mittag-Leffler functions

Let $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$. The Mittag-Leffler function $E_{\alpha,\beta}$ is defined for $z \in \mathbb{C}$ by

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

For $(\alpha, \lambda) \in (0,1) \times \mathbb{R}_+$, we also define

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^\alpha), \quad t > 0,$$

$$F^{\alpha,\lambda} = \int_0^t f^{\alpha,\lambda}(s)ds, \quad t \geq 0.$$

The function $f^{\alpha,\lambda}$ is a density function on $\mathbb{R}_+$ called Mittag-Leffler density function.
For $\alpha \in (1/2, 1)$, $f^{\alpha,\lambda}$ is square-integrable and its Laplace transform is given for $z \geq 0$ by

$$\hat{f}^{\alpha,\lambda}(z) = \int_0^\infty f_{\alpha,\lambda}(s) e^{-zs} ds = \frac{\lambda}{\lambda + z^\alpha}.$$

Finally, we can show that

$$I^{1-\alpha} f^{\alpha,\lambda}(t) = \lambda\big(1 - F^{\alpha,\lambda}(t)\big).$$

Further properties of $f^{\alpha,\lambda}$ and $F^{\alpha,\lambda}$ can be found in [HMS11, Mai, MH08].

Part II

Pricing under the rough Heston model

CHAPTER II

Characteristic function of rough Heston models

Abstract

It has been recently shown that rough volatility models, where the volatility is driven by a fractional Brownian motion with small Hurst parameter, provide very relevant dynamics in order to reproduce the behavior of both historical and implied volatilities. However, due to the non-Markovian nature of the fractional Brownian motion, they raise new issues when it comes to derivatives pricing. Using an original link between nearly unstable Hawkes processes and fractional volatility models, we compute the characteristic function of the log-price in rough Heston models. In the classical Heston model, the characteristic function is expressed in terms of the solution of a Riccati equation. Here we show that rough Heston models exhibit quite a similar structure, the Riccati equation being replaced by a fractional Riccati equation.

Keywords: Rough volatility models, rough Heston models, Hawkes processes, fractional Brownian motion, fractional Riccati equation, limit theorems.

1 Introduction

The celebrated Heston model is a one-dimensional stochastic volatility model where the asset price S follows the following dynamic:

$$dS_t = S_t\sqrt{V_t}dW_t$$

$$dV_t = \lambda(\theta - V_t)dt + \lambda\nu\sqrt{V_t}dB_t. \tag{1}$$

Here the parameters λ, θ, V_0 and ν are positive, and W and B are two Brownian motions with correlation coefficient ρ, that is $\langle dW_t, dB_t\rangle = \rho dt$.

The popularity of this model is probably due to three main reasons:

- It reproduces well several important stylized facts of low frequency price data, namely leverage effect, time-varying volatility and fat tails, see [BP03, Chr82, DY02, Man97].

- It generates very reasonable shapes and dynamics for the implied volatility surface. Indeed, the "volatility of volatility" parameter ν enables us to control the smile, the correlation parameter ρ to deal with the skew, and the initial volatility V_0 to fix the at-the-money volatility level, see [FJL12, Gat11, JKWW11, Poo09]. Furthermore, as observed in markets and in contrast to local volatility models, in Heston model, the volatility smile moves in the same direction as the underlying and the forward smile does not flatten with time, see [Gat11, JR13, JR16a, MP15b].

- There is an explicit formula for the characteristic function of the asset log-price, see [Hes93]. From this formula, efficient numerical methods have been developed, allowing for instantaneous model calibration and pricing of derivatives, see [AMST07, CM99, KJ05, Lew01].

In the classical Heston model, the volatility follows a Brownian semi-martingale. However, it is demonstrated in [GJR18] that for a very wide range of assets, historical volatility time-series exhibit a behavior which is much rougher than that of a Brownian motion. More precisely, the dynamic of log-volatility is typically very well modeled by a fractional Brownian motion with Hurst parameter of order 0.1. For example, it is shown in [GJR18] that in practice, the empirical moment of order $q>0$ of log-volatility increments

$$\log(V_{t+\Delta})-\log(V_t),$$

is proportional to Δ^{qH} with H of order 0.1, and this for any reasonable scale of interest Δ (from Δ equal to one day to hundreds of days). This corresponds to a rough fractional dynamic with Hurst parameter $H=0.1$. Beyond moments, it is also established in [GJR18] that the empirical correlation structure of volatility is very well reproduced when using rough fractional volatility models. These findings have been confirmed by further studies, see [BLP16, LMPR18]. Moreover, considering a fractional Brownian motion with small Hurst parameter also enables us to obtain remarkable fits for the whole volatility surface. In particular, contrary to most stochastic volatility models, rough volatility models generate an exploding term structure for the at-the-money skew when maturity goes to zero, which is very commonly observed in practice, see [BFG16, GJR18] and Section 5. Finally, convincing microstructural foundations for rough volatility models are provided in [JR16b] and Chapter I, see also Section 2.

Hence, in this paper, we are interested in the fractional versions of the Heston model. Our main goal is to design an efficient pricing methodology for such models, in the spirit of the one introduced by Heston in the classical case. This is particularly important in fractional volatility models where the use of Monte-Carlo methods can be quite intricate due to the non-Markovian nature of the fractional Brownian motion, see [BLP17].

We now define our so-called rough Heston model. Let us recall that a fractional Brownian motion W^H with Hurst parameter $H\in(0,1)$ can be built through the Mandelbrot-van Ness representation

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

The kernel $(t-s)^{H-\frac{1}{2}}$ in (2) plays a central role in the rough dynamic of the fractional Brownian motion for $H<1/2$. In particular, one can show that the process

$$\int_0^t (t-s)^{H-\frac{1}{2}} dW_s$$

has Hölder regularity $H-\varepsilon$ for any $\varepsilon>0$. In order to allow for a rough behavior of the volatility in a Heston-type model, we naturally introduce the kernel $(t-s)^{\alpha-1}$ in a Heston-like stochastic volatility process as follows

$$dS_t = S_t\sqrt{V_t}dW_t$$
$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta-V_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda\nu\sqrt{V_s}dB_s. \tag{3}$$

The parameters λ, θ, V_0 and ν in (3) are positive and play the same role as in (1), and here also W and B are two Brownian motions with correlation ρ. The additional parameter α belongs to $(1/2,1)$ and governs the smoothness of the volatility sample paths. More precisely, we show in this paper that the model is well-defined and that the volatility trajectories have almost surely Hölder regularity $\alpha-1/2-\varepsilon$, for any $\varepsilon>0$. When $\alpha=1$, Models (3) and (1) coincide, and we retrieve the classical Heston model. Therefore it is natural to view (3) as a rough version of the Heston model and to call it rough Heston model. In term of Hurst parameter H, $\alpha=H+1/2$. Nevertheless, note that other definitions of rough Heston models can make sense, see [GJR14] for an alternative definition and some asymptotic results.

Our aim in this work is to derive a Heston-type formula for the characteristic function of the log-price in Model (3). In the classical case ($\alpha=1$, Model (1)), this formula is proved in [Hes93]. It is obtained using the fact that Model (1) is Markovian and time-homogeneous, and applying Itô's formula to the function

$$L(t,a,V_t,S_t) = \mathbb{E}[e^{ia\log(S_T)}|\mathscr{F}_t], \quad \mathscr{F}_t = \sigma(W_s,B_s; s\le t), \quad a\in\mathbb{R}.$$

The process L being a martingale, the following Feynman-Kac partial differential equation for L is easily obtained

$$-\partial_t L(t,a,S,V) = \big(\lambda(\theta-V)\partial_v + \frac{1}{2}(\lambda\nu)^2 V\partial^2_{vv} + \frac{1}{2}S^2V\partial^2_{ss} + \rho\nu\lambda SV\partial^2_{sv}\big)L(t,a,S,V),$$

with boundary condition $L(T,a,S,V)=e^{ia\log(S)}$. From this PDE, it can be checked that the characteristic function of the log-price $X_t=\log(S_t/S_0)$ satisfies

$$\mathbb{E}[e^{iaX_t}] = \exp\big(g(a,t)+V_0h(a,t)\big),$$

where h is solution of the following Riccati equation

$$\partial_t h(a,t) = \frac{1}{2}(-a^2-ia) + \lambda(ia\rho\nu-1)h(a,t) + \frac{(\lambda\nu)^2}{2}h^2(a,t), \quad h(a,0)=0, \tag{4}$$

and

$$g(a,t) = \theta\lambda\int_0^t h(a,s)ds.$$

Solving this Riccati equation leads to the closed-form formula for the characteristic function of the log-price given in [Hes93].

In the case $\alpha < 1$, the rough Heston model (3) is not Markovian and the variance process is no longer a semi-martingale. Hence the strategy initially used by Heston and presented above seems very hard to adapt to our setting. Here we resort to a completely different and original approach based on point processes. Indeed, our methodology finds its root in [JR16b] and Chapter I which provide microstructural foundations for rough volatility models. In these papers, it is shown that some well-designed microstructure models, reproducing the stylized facts of modern financial markets at high frequency, give rise in the long run to rough volatility models. These microstructure models, that we describe in more details in Section 2, are based on so-called nearly unstable Hawkes processes. In this paper, inspired by these results and using again Hawkes processes, we design a suitable sequence of point processes which converges to Model (3). Exploiting the specific structure of our point processes, we derive their characteristic function, which leads us in the limit to that of the log-price in the rough Heston model (3).

Our main result is that, quite surprisingly, the characteristic function of the log-price in rough Heston models exhibits the same structure as the one obtained in the classical Heston model. The difference is that the Riccati equation (4) is replaced by a fractional Riccati equation, where a fractional derivative appears instead of a classical derivative. More precisely, we obtain

$$\mathbb{E}[e^{iaX_t}] = \exp\big(g_1(a,t) + V_0 g_2(a,t)\big),$$

where

$$g_1(a,t) = \theta\lambda \int_0^t h(a,s)ds, \quad g_2(a,t) = I^{1-\alpha}h(a,t),$$

and $h(a,.)$ is a solution of the following fractional Riccati equation

$$D^\alpha h(a,t) = \frac{1}{2}(-a^2 - ia) + \lambda(ia\rho\nu - 1)h(a,t) + \frac{(\lambda\nu)^2}{2}h^2(a,t), \quad I^{1-\alpha}h(a,0) = 0,$$

with D^α and $I^{1-\alpha}$ the fractional derivative and integral operators defined in (18) and (19). Remark that when $\alpha = 1$, this result indeed coincides with the classical Heston's result. However, note that for $\alpha < 1$, the solutions of such Riccati equations are no longer explicit. Nevertheless, they are easily solved numerically, see Section 5.

The paper is organized as follows. In Section 2, we build a sequence of Hawkes-type processes which converges to the rough Heston model (3). Then we study in Section 3 the characteristic function of these processes and show in Section 4 that it enables us to derive the characteristic function of the log-price in Model (3). Numerical illustrations are given in Section 5 and some proofs are relegated to Section 6. Finally, some useful technical results are given in an appendix.

2 From Hawkes processes to rough Heston models

We build in this section a sequence of Hawkes-type processes which converges to the rough Heston model (3). This construction is inspired by Chapter I. In this work, microstructural foundations for rough Heston models are provided. This is done designing suitable sequences of ultra high frequency price models which reproduce the stylized facts of modern markets microstructure and converge in the long run to rough Heston models. These microscopic price models are based on Hawkes processes. So that the reader can understand the genesis of our original methodology to compute the characteristic function in rough Heston models, we recall here the main ideas and results in Chapter I.

2.1 Microstructural foundations for rough Heston models

In Chapter I, we consider a sequence of bi-dimensional Hawkes processes $(N^{T,+}, N^{T,-})$ indexed by $T > 0$ going to infinity[1] and with intensity[2]

$$\lambda_t^T = \begin{pmatrix} \lambda_t^{T,+} \\ \lambda_t^{T,-} \end{pmatrix} = \mu_T \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \int_0^t a_T \phi(t-s).\begin{pmatrix} dN_s^{T,+} \\ dN_s^{T,-} \end{pmatrix}, \tag{5}$$

with

$$\phi = \begin{pmatrix} \varphi_1 & \varphi_3 \\ \varphi_2 & \varphi_4 \end{pmatrix}.$$

Here the φ_i are measurable non-negative deterministic functions and μ_T and $0 < a_T < 1$ are some deterministic sequences of positive real numbers, see [BDHM13b] and the references therein for more details about the definition of Hawkes processes. Then in Chapter I, inspired by [BDHM13a, BDHM13b, JR15], we consider the following ultra high frequency tick-by-tick model for the transaction price P_t^T,

$$P_t^T = N_t^{T,+} - N_t^{T,-}. \tag{6}$$

Hence $N_t^{T,+}$ represents the number of upward jumps of one tick of the transaction price over the period $[0, t]$ and $N_t^{T,-}$ the number of downward jumps. The relevance of this Hawkes-based modeling is that it enables us to encode very easily the most important stylized facts of high frequency markets in term of the parameters of the Hawkes process. We now give these stylized facts and their translation in terms of the model parameters, referring to Chapter I for more details.

- Markets are highly endogenous: In the high frequency trading context, most orders have no real economic motivation. They are rather sent by algorithms as reactions to other orders. In the Hawkes framework, this amounts to work with so-called *nearly unstable Hawkes processes.* This means that the stability condition

$$\mathcal{S}\Big(\int_0^\infty a_T \phi(s) ds\Big) < 1,$$

[1]Of course by T we implicitly mean T_n with $n \in \mathbb{N}$ tending to infinity.

[2]From now on we write a dot between quantities to emphasize matrix product.

where $\mathscr{S}$ denotes the spectral radius operator, should almost be saturated and that the intensity of exogenous orders, namely μ_T, should be small, see [HBB13, JR16b, JR15] and Chapter I. In term of model parameters, suitable constraints are therefore

$$a_T \to 1, \quad \mathscr{S}\Big(\int_0^\infty \phi(s)ds\Big) = 1, \quad \mu_T \to 0.$$

- It is not an easy task to make money with high frequency strategies on highly liquid electronic markets. Hence some “no statistical arbitrage" mechanisms should be in force. We translate this assuming that in the long run, there are on average as many upward as downward jumps. This corresponds to the assumption

$$\varphi_1 + \varphi_3 = \varphi_2 + \varphi_4.$$

- Buying is not the same action as selling. This means that buy market orders and sell market orders are not symmetric orders. To see this, consider for example a market maker, with an inventory which is typically positive. After each order he receives, he modifies his bid and ask quotes, reflecting the market impact of the received order. This means that after a buy order, he will increase his ask quote and he will decrease his bid quote after a sell order. However, he typically raises the price by less following a buy order than he lowers the price following the same size sell order. Indeed, inventory becomes smaller after a buy order, which is a good thing for him, whereas it increases after a sell order. This creates a liquidity asymmetry on the bid and ask sides of the order book. This can be modeled in the Hawkes framework assuming that

$$\varphi_3 = \beta\varphi_2,$$

for some $\beta > 1$. Hence, the matrix ϕ finally takes the form

$$\phi = \begin{pmatrix} \varphi_1 & \beta\varphi_2 \\ \varphi_2 & \varphi_1 + (\beta-1)\varphi_2 \end{pmatrix}.$$

- A significant amount of transactions is part of metaorders, which are large orders whose execution is split in time by trading algorithms. This is translated into a heavy tail assumption on the functions φ_1 and φ_2, namely that there exists $1/2 < \alpha < 1$ (typically around 0.6 in practice, see [BJM16, HBB13]) and $C > 0$ such that

$$\alpha x^\alpha \int_x^\infty \varphi_1(s) + \beta\varphi_2(s)ds \underset{x\to\infty}{\to} C.$$

Furthermore, it is shown in [JR16b] that for a given α, there is only one way to make μ_T tends to zero and a_T tends to one so that the limit of the price is not degenerate. More precisely,

$$(1-a_T)T^\alpha \underset{T\to\infty}{\to} \lambda^*, \quad \mu_T T^{1-\alpha} \underset{T\to\infty}{\to} \mu,$$

for some positive λ^* and μ.

Under the above assumptions, it is proved in Chapter I that the properly rescaled microscopic price process

$$\sqrt{\frac{1-a_T}{\mu T^{\alpha}}}P^T_{tT}, \quad t \in [0,1],$$

where P^T is defined in (11), converges in law as T tends to infinity to the following macroscopic price dynamic P,

$$P_t = \frac{\sqrt{2}}{1-\int_0^{\infty}(\varphi_1-\varphi_2)}\int_0^t \sigma_s dW_s,$$

$$\sigma_t^2 = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(1-\sigma_s^2)ds + \frac{1}{\Gamma(\alpha)}\lambda\nu\int_0^t (t-s)^{\alpha-1}\sigma_s dB_s, \tag{7}$$

where (W,B) is a bi-dimensional correlated Brownian motion with correlation

$$\rho = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}$$

and

$$\nu = \sqrt{\frac{2(1+\beta^2)}{\lambda^*\mu(1+\beta)^2}}, \quad \lambda = \lambda^*\frac{\alpha}{C\Gamma(1-\alpha)}.$$

Hence this result shows that the main stylized facts of modern electronic markets naturally give rise to a very rough behavior of the volatility. Indeed, recall that the Hurst parameter corresponds to $\alpha - 1/2$.

Inspired by this result, our idea is to study the characteristic function of some kind of microscopic price processes in order to deduce that of our rough Heston macroscopic price of interest (3). However, the developments presented above cannot be directly applied and need to be adapted. Indeed, remark that in (7), $\sigma_0 = 0$. This does not correspond to the case of (3), where having a non-zero initial volatility is of course crucial for the model to be relevant in practice. Thus we need to modify the sequence of Hawkes-type processes to obtain a non-degenerate initial volatility in the limit. This is actually a non-trivial issue. However, this can be achieved replacing μ_T in (5) by an inhomogeneous Poisson intensity $\hat{\mu}_T(t)$. We explain how such $\hat{\mu}_T(t)$ can be found in the next section.

2.2 The role of the Poisson rate

We work on a sequence of probability spaces $(\Omega^T, \mathscr{F}^T, \mathbb{P}^T)$, indexed by $T > 0$ (going to infinity), on which $N^T = (N^{T,+}, N^{T,-})$ is a bi-dimensional Hawkes process with intensity

$$\lambda_t^T = \begin{pmatrix}\lambda_t^{T,+}\\ \lambda_t^{T,-}\end{pmatrix} = \hat{\mu}_T(t)\begin{pmatrix}1\\1\end{pmatrix} + \int_0^t \phi^T(t-s).dN_s^T. \tag{8}$$

For a given T, the probability space is equipped with the filtration $(\mathscr{F}_t^T)_{t\geq 0}$, where $\mathscr{F}_t^T$ is the σ-algebra generated by $(N_s^T)_{s\leq t}$. Because our goal is to design a sequence of processes leading in the limit to a rough Heston dynamic, we consider the same kind of assumptions on the matrix ϕ^T as those described in the previous section. However, here we can be very specific because we just need to find one convenient sequence of processes. That is why we make a particular choice for the heavy-tailed functions defining ϕ^T, using Mittag-Leffler functions, see Section II.A in the Appendix for definition and some properties. Indeed, these functions are very convenient in order to carry out computations. More precisely, our assumptions on ϕ^T are as follows.

Assumption 1. *There exist $\beta \geq 0$, $1/2 < \alpha < 1$ and $\lambda > 0$ such that*

$$a_T = 1 - \lambda T^{-\alpha}, \quad \phi^T = \varphi^T \chi,$$

where

$$\chi = \frac{1}{\beta+1}\begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix}, \quad \varphi^T = a_T \varphi, \quad \varphi = f^{\alpha,1},$$

with $f^{\alpha,1}$ the Mittag-Leffler density function defined in Appendix.

Remark 1. *From Appendix II.A, we get that we are working in the nearly unstable heavy tail case because*

$$\int_0^\infty \varphi(s)ds = 1$$

and

$$\alpha x^\alpha \int_x^\infty \varphi(t)dt \underset{x\to\infty}{\longrightarrow} \frac{\alpha}{\Gamma(1-\alpha)}.$$

We now give intuitions about the need to use a non-constant Poisson intensity $\hat{\mu}_T(t)$. First, note that under Assumption 1,

$$\lambda_t^{T,+} = \lambda_t^{T,-}.$$

The asymptotic behavior of the renormalized intensity processes $\lambda_t^{T,+}$ and $\lambda_t^{T,-}$ will give us that of the volatility in our limiting macroscopic price model. Thus, we need to understand the long term limit of $\lambda_t^{T,+}$. Let us write

$$M_t^T = (M_t^{T,+}, M_t^{T,-}) = N_t^T - \int_0^t \lambda_s^T ds$$

for the martingale associated to the point process N_t^T. We easily obtain

$$\lambda_t^{T,+} = \hat{\mu}_T(t) + \int_0^t \varphi^T(t-s)\lambda_s^{T,+}ds + \frac{1}{1+\beta}\int_0^t \varphi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}).$$

Now let

$$\psi^T = \sum_{k\geq 1} (\varphi^T)^{*k},$$

where $(\varphi^T)^{*1} = \varphi^T$ and for $k > 1$, $(\varphi^T)^{*k}(t) = \int_0^t \varphi^T(s)(\varphi^T)^{*(k-1)}(t-s)ds$. Using Lemma 1 in the Appendix together with Fubini theorem and the fact that $\psi^T * \varphi^T = \psi^T - \varphi^T$, we get

$$\lambda_t^{T,+} = \hat{\mu}_T(t) + \int_0^t \psi^T(t-s)\hat{\mu}_T(s)ds + \frac{1}{1+\beta}\int_0^t \psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}). \tag{9}$$

Following Chapter I, the inhomogeneous intensity $\hat{\mu}_T(t)$ should be of order μ_T with

$$\mu_T = \mu T^{\alpha-1},$$

where μ is some positive constant. In Chapter I, it is shown that the right normalization for the intensity in order to get a non-degenerate limit, is to consider $(1-a_T)\lambda_{tT}^{T,+}/\mu_T$. The same applies here and thus we define the renormalized intensity

$$C_t^T = \frac{1-a_T}{\mu_T}\lambda_{tT}^{T,+}.$$

After obvious computations, this can be written

$$C_t^T = \frac{1-a_T}{\mu_T}\hat{\mu}_T(tT)+\int_0^t T(1-a_T)\psi^T\big(T(t-s)\big)\frac{\hat{\mu}_T(Ts)}{\mu_T}ds+\nu\int_0^t T(1-a_T)\psi^T\big(T(t-s)\big)\sqrt{C_s^T}dB_s^T,$$

where

$$B_t^T = \int_0^{tT}\frac{dM_s^{T,+}+\beta dM_s^{T,-}}{\sqrt{T(\lambda_s^{T,+}+\beta^2\lambda_s^{T,-})}},\quad \nu = \sqrt{\frac{1+\beta^2}{\lambda\mu(1+\beta)^2}}.$$

Using the fact that the Laplace transform $\hat{f}^{\alpha,\lambda}$ of the Mittag-Leffler density function $f^{\alpha,\lambda}$ is given by

$$\hat{f}^{\alpha,\lambda}(z) = \frac{\lambda}{\lambda+z^\alpha},$$

we easily obtain that

$$(1-a_T)T\psi^T(T.) = a_T f^{\alpha,\lambda}, \tag{10}$$

see Section II.A in Appendix. This leads to the following expression for C^T:

$$C_t^T = \frac{1-a_T}{\mu_T}\hat{\mu}_T(tT)+\int_0^t a_T f^{\alpha,\lambda}(t-s)\frac{\hat{\mu}_T(Ts)}{\mu_T}ds+\nu\int_0^t a_T f^{\alpha,\lambda}(t-s)\sqrt{C_s^T}dB_s^T.$$

Computing the quadratic variation of B^T, it is easy to see that it converges to a Brownian motion B. Now, if as in Chapter I we take $\hat{\mu}_T(t) = \mu_T$, we obtain that C^T should then give in the limit a process with starting value equal to zero. Nevertheless, we also get the intuition that a non-constant $\hat{\mu}_T$ can lead to a non-trivial initial value. From the computations in the proof of Theorem 1, it will become clear that the right choice of $\hat{\mu}_T$ is as follows

Assumption 2. *The baseline intensity $\hat{\mu}_T$ is given by*

$$\hat{\mu}_T(t) = \mu_T+\xi\mu_T\Big(\frac{1}{1-a_T}(1-\int_0^t\varphi^T(s)ds)-\int_0^t\varphi^T(s)ds\Big),$$

with $\xi>0$ and $\mu_T = \mu T^{\alpha-1}$ for some $\mu>0$.

Remark 2. *Note that $\hat{\mu}_T$ can also be written as follows*

$$\hat{\mu}_T(t) = \mu_T+\xi\mu_T\Big(\frac{T^\alpha}{\lambda}\int_t^\infty\varphi(s)ds+\lambda T^{-\alpha}\int_0^t\varphi(s)ds\Big).$$

This shows that $\hat{\mu}_T$ is a positive function and thus that the intensity process λ_t^T in (8) *is well-defined.*

2.3 The rough limits of Hawkes processes

We now give a rigorous statement about the limiting behavior of our specific sequence of bi-dimensional nearly unstable Hawkes processes with heavy tails. For $t \in [0,1]$, we define

$$X_t^T = \frac{1-a_T}{T^\alpha \mu} N_{tT}^T, \quad \Lambda_t^T = \frac{1-a_T}{T^\alpha \mu} \int_0^{tT} \lambda_s^T ds, \quad Z_t^T = \sqrt{\frac{T^\alpha \mu}{1-a_T}} (X_t^T - \Lambda_t^T).$$

Using a similar approach as that in Chapter I, we obtain the following result whose proof is given in Section 6.

Theorem 1. *As $T \to \infty$, under Assumptions 1 and 2, the process $\left(\Lambda_t^T, X_t^T, Z_t^T\right)_{t\in[0,1]}$ converges in law for the Skorokhod topology to (Λ, X, Z) where*

$$\Lambda_t = X_t = \int_0^t Y_s ds \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad Z_t = \int_0^t \sqrt{Y_s} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix},$$

and Y is the unique solution of the rough stochastic differential equation[3]

$$Y_t = \xi + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda(1-Y_s) ds + \lambda \sqrt{\frac{1+\beta^2}{\lambda\mu(1+\beta^2)}} \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{Y_s} dB_s,$$

where

$$B = \frac{B^1 + \beta B^2}{\sqrt{1+\beta^2}}$$

and (B^1, B^2) is a bi-dimensional Brownian motion. Furthermore, for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

Hence Theorem 1 shows that designing our sequence of bi-dimensional Hawkes processes in a suitable way, its limit is differentiable and its derivative exhibits a rough Cox-Ingersoll-Ross like behavior, with non-zero initial value. This is exactly what we need for the limiting volatility of our microscopic price processes. Indeed, thanks to Theorem 1, we are now able to build such microscopic processes converging to the log-price in (3). More precisely, for $\theta > 0$, let us define

$$P^T = \sqrt{\frac{\theta}{2}} \sqrt{\frac{1-a_T}{T^\alpha \mu}} (N_{.T}^{T,+} - N_{.T}^{T,-}) - \frac{\theta}{2} \frac{1-a_T}{T^\alpha \mu} N_{.T}^{T,+} = \sqrt{\frac{\theta}{2}} (Z^{T,+} - Z^{T,-}) - \frac{\theta}{2} X^{T,+}. \tag{11}$$

We have the following corollary of Theorem 1.

Corollary 1. *As $T \to \infty$, under Assumptions 1 and 2, the sequence of processes $(P_t^T)_{t\in[0,1]}$ converges in law for the Skorokhod topology to*

$$P_t = \int_0^t \sqrt{V_s} dW_s - \frac{1}{2} \int_0^t V_s ds,$$

[3] Note that we call this equation rough because of the presence of the kernel $(t-s)^{\alpha-1}$. However, it is not directly related to rough paths theory.

where V is the unique solution of the rough stochastic differential equation

$$V_t = \theta\xi + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta - V_s)ds + \lambda\sqrt{\frac{\theta(1+\beta^2)}{\lambda\mu(1+\beta)^2}}\frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{V_s}dB_s,$$

with (W,B) a correlated bi-dimensional Brownian motion whose bracket satisfies

$$d\langle W,B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

Thus, we have succeeded in building a sequence of microscopic processes P^T, defined by (11), which converges to (the logarithm of) our rough Heston process of interest (3). Now our goal is to use the result of Corollary 1 to compute the characteristic function of the log-price in the rough Heston model (3). This is done in the next two sections.

3 The characteristic function of multivariate Hawkes processes

We have seen in the previous section that our sequence of Hawkes-based microscopic price processes converges to the log-price in the rough Heston model (3). Therefore, if we are able to compute the characteristic function for the microscopic price, its limit will give us that of the log-price in a rough Heston model. We actually provide a more general result here, deriving the characteristic function of a multivariate Hawkes process (recall that a bi-dimensional Hawkes process is the building block for our microscopic price process (11)). Hence we extend here some results already proved in [HO74] in the one-dimensional case.

3.1 Cluster-based representation

To derive our characteristic function, the representation of Hawkes processes in term of clusters, see [HO74], is very useful. We recall it now. Let us consider a d-dimensional Hawkes process $N = (N^1, ..., N^d)$ with intensity

$$\lambda_t = \begin{pmatrix}\lambda_t^1\\ \vdots \\ \lambda_t^d\end{pmatrix} = \mu(t) + \int_0^t \phi(t-s).dN_s, \tag{12}$$

where $\mu : \mathbb{R}_+ \to \mathbb{R}_+^d$ is locally integrable and $\phi : \mathbb{R}_+ \to \mathcal{M}^{\mathbf{d}}(\mathbb{R}_+)$ has integrable components such that

$$\mathcal{S}\Big(\int_0^\infty \phi(s)ds\Big) < 1.$$

The law of such process can be described through a population approach. Consider that there are d types of individuals and for a given type, an individual can be either a migrant or the descendant of a migrant. Then the dynamic goes as follows from time $t = 0$:

- Migrants of type $k \in \{1,..,d\}$ arrive as a non-homogenous Poisson process with rate $\mu_k(t)$.

- Each migrant of type $k \in \{1,..,d\}$ gives birth to children of type $j \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{j,k}(t)$.
- Each child of type $k \in \{1,..,d\}$ also gives birth to other children of type $j \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{j,k}(t)$.

Then, for $k \in \{1,..,d\}$, N_t^k can be taken as the number up to time t of migrants and children born with type k. Indeed, the population approach above and the theoretical characterization (12) define the same point process law.

3.2 The result

Let $L(a,t)$ be the characteristic function of the Hawkes process N,

$$L(a,t) = \mathbb{E}[\exp(ia.N_t)], \quad t \geq 0, \quad a \in \mathbb{R}^d,$$

where $a.N_t$ stands for the scalar product of a and N_t. The cluster-based representation of multivariate Hawkes processes enables us to show the following result, proved in Section 3.3, for their characteristic function.

Theorem 2. *We have*

$$L(a,t) = \exp\Big(\int_0^t \big(C(a,t-s) - \mathbf{1}\big).\mu(s)ds\Big),$$

where $C : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{C}^d$ *is solution of the following integral equation:*

$$C(a,t) = \exp\Big(ia + \int_0^t \phi^*(s).(C(a,t-s) - \mathbf{1})ds\Big),$$

with $\phi^*(s)$ *the transpose of* $\phi(s)$.

From Theorem 2, we are able to derive in Section 4 the characteristic function of rough Heston models.

3.3 Proof of Theorem 2

We now give the proof of Theorem 2, exploiting the population construction presented in Section 3.1. We start by defining d auxiliary independent d-dimensional point processes $(\tilde{N}^{k,j})_{1 \leq j \leq d}$, $k \in \{1,...,d\}$, defined as follows for each given $k \in \{1,..,d\}$:

- Migrants of type $j \in \{1,...,d\}$ arrive as a non-homogenous Poisson process with rate $\phi_{j,k}(t)$.
- Each migrant of type $j \in \{1,..,d\}$ gives birth to children of type $l \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{l,j}(t)$.
- Each child of type $j \in \{1,..,d\}$ also gives birth to other children of type $l \in \{1,..,d\}$ following a non-homogenous Poisson process with rate $\phi_{l,j}(t)$.

For a given $k \in \{1,..,d\}$, $\tilde{N}_t^{k,j}$ corresponds to the number, up to time t, of migrants and children with type j. A simple but crucial remark is that $(\tilde{N}^{k,j})_{1\le j\le d}$ is actually also a multivariate Hawkes process with migrant rate $(\phi_{j,k})_{1\le j\le d}$ and kernel matrix ϕ. We write $L_k(a,t)$ for its characteristic function

$$L_k(a,t) = \mathbb{E}\big[\exp(ia.(\tilde{N}_t^{k,j})_{1\le j\le d})\big], \quad t \ge 0, \quad a \in \mathbb{R}^d.$$

Now let us come back to the initial Hawkes process of interest N defined by (12). For each $k \in \{1,...,d\}$ and $t \ge 0$, let $N_t^{0,k}$ be the number of its migrants of type k arrived up to time t. Recall that the $N^{0,k}$, $1 \le k \le d$, are independent Poisson processes with rates $\mu_k(t)$. We also define $T_1^k < ... < T^k_{N_t^{0,k}} \in [0,t]$ the arrival times of migrants of type k of the Hawkes process N, up to time t. Using the population approach presented in Section 3.1, it is clear that at time t, the number of descendants of different types of a migrant of type k arrived at time T_u^k has the same law as $(\tilde{N}^{k,j}_{t-T_u^k})_{1\le j\le d}$, where $\tilde{N}$ is taken independent from N. Consequently,

$$N_t^k \underset{law}{=} N_t^{0,k} + \sum_{1\le j\le d} \sum_{1\le l\le N_t^{0,j}} \tilde{N}^{j,k,(l)}_{t-T_l^j}, \tag{13}$$

where the $(\tilde{N}^{j,k,(l)})_{1\le k\le d}$, $1 \le j \le d$, $l \in \mathbb{N}$ are independent copies of $(\tilde{N}^{j,k})_{1\le k\le d}$, $1 \le j \le d$, also independent of $N^0 = (N^{0,k})_{1\le k\le d}$.

From (13), we derive that conditional on N^0,

$$\begin{aligned}\mathbb{E}\big[\exp(ia.N_t)|N^0\big] &= \exp(ia.N_t^0) \prod_{1\le j\le d} \prod_{1\le l\le N_t^{0,j}} \mathbb{E}\big[\exp(ia.(\tilde{N}^{j,k,(l)}_{t-T_l^j})_{1\le k\le d}|N^0)\big] \\ &= \exp(ia.N_t^0) \prod_{1\le j\le d} \prod_{1\le l\le N_t^{0,j}} L_j(a, t-T_l^j).\end{aligned}$$

Now, for a given $k \in \{1,...,d\}$, conditional on $N_t^{0,k}$, it is well-known that $(T_1^k,...,T^k_{N_t^{0,k}})$ has the same law as $(X_{(1)},...,X_{(N_t^{0,k})})$ the order statistics built from iid variables $(X_1,..,X_{N_t^{0,k}})$ with density $\frac{\mu_k(s)1_{s\le t}}{\int_0^t \mu_k(s)ds}$. Thus we get

$$\mathbb{E}\big[\exp(ia.N_t)|N_t^0\big] = \exp(ia.N_t^0) \prod_{1\le j\le d} \Big(\int_0^t L_j(a,t-s)\frac{\mu_j(s)}{\int_0^t \mu_j(s)ds}ds\Big)^{N_t^{0,j}}.$$

Therefore,

$$L(a,t) = \prod_{1\le j\le d} \exp\Big(\Big(\int_0^t e^{ia_j} L_j(a,t-s)\frac{\mu_j(s)}{\int_0^t \mu_j(s)ds}ds - 1\Big)\int_0^t \mu_j(s)ds\Big).$$

Thus we finally obtain

$$L(a,t) = \exp\Big(\sum_{1\le j\le d} \int_0^t (e^{ia_j} L_j(a,t-s) - 1)\mu_j(s)ds\Big). \tag{14}$$

In the same way, because $(\tilde{N}^{k,j})_{1\leq j\leq d}$ is a multivariate Hawkes process with migrant rate $(\phi_{j,k})_{1\leq j\leq d}$ and kernel matrix ϕ, we get

$$L_k(a,t) = \exp\Big(\sum_{1\leq j\leq d}\int_0^t (e^{ia_j}L_j(a,t-s)-1)\phi_{j,k}(s)ds\Big). \tag{15}$$

Let us define

$$C(a,t) = \big(e^{ia_j}L_j(a,t)\big)_{1\leq j\leq d}.$$

From (14), we have that

$$L(a,t) = \exp\Big(\int_0^t (C(a,t-s)-\mathbf{1}).\mu(s)ds\Big)$$

and from (15), we deduce that C is solution of the following integral equation

$$C(a,t) = \exp\big(ia + \int_0^t \phi^*(s).(C(a,t-s)-\mathbf{1})ds\big).$$

This ends the proof of Theorem 2.

4 The characteristic function of rough Heston models

We give in this section our main theorem, that is the characteristic function for the log-price in rough Heston models (3). It is obtained combining the convergence result for Hawkes processes stated in Corollary 1 together with the characteristic function for multivariate Hawkes processes derived in Theorem 2. We start with some intuitions about the result.

4.1 Intuition about the result

We consider the rough Heston model (3). The parameters of the dynamic in (3) are here given in term of those of the sequence of processes P^T defined in (11). More precisely, we set

$$V_0 = \xi\theta, \quad \rho = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}, \quad \nu = \sqrt{\frac{\theta(1+\beta^2)}{\lambda\mu(1+\beta)^2}},$$

and λ and θ are the same as those in the dynamic of P^T. Remark that the fact that $\beta \geq 0$ implies that $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$[4].. We also write $P_t = \log(S_t/S_0)$. From Corollary 1, we know that

$$P^T = \sqrt{\frac{\lambda\theta}{2\mu}}T^{-\alpha}(N^{T,+}_{.T} - N^{T,-}_{.T}) - \frac{\lambda\theta}{2\mu}T^{-2\alpha}N^{T,+}_{.T}$$

converges in law to P as T tends to infinity, where $N^T = (N^{T,+}, N^{T,-})$ is a sequence of bi-dimensional Hawkes processes satisfying Assumptions 1 and 2. Let us write $L^T((a,b),u)$ for the

[4]Actually, using a more complex framework of Hawkes processes, one can show that the results still hold for any $\rho \in [-1,1]$, see Chapter III

characteristic function of the process N^T at time u at point (a,b) and L_p for the characteristic function of P. The convergence in law implies that of $L^T((a_T^+,a_T^-),tT)$ towards $L_p(a,t)$, where

$$a_T^+ = a\sqrt{\frac{\lambda\theta}{2\mu}}T^{-\alpha} - a\frac{\lambda\theta}{2\mu}T^{-2\alpha}, \quad a_T^- = -a\sqrt{\frac{\lambda\theta}{2\mu}}T^{-\alpha}.$$

From Theorem 2, we know that

$$L^T\big((a_T^+,a_T^-),tT\big) = \exp\Big(\int_0^{tT}\hat{\mu}_T(s)\big((C^{T,+}((a_T^+,a_T^-),tT-s)-1)+(C^{T,-}((a_T^+,a_T^-),tT-s)-1)\big)ds\Big),$$

where $C^T((a_T^+,a_T^-),t) = \big(C^{T,+}((a_T^+,a_T^-),t),C^{T,-}((a_T^+,a_T^-),t)\big) \in \mathscr{M}^{\mathbf{1\times 2}}(\mathbb{C})$ is solution of

$$C^T\big((a_T^+,a_T^-),t\big) = \exp\big(i(a_T^+,a_T^-) + \int_0^t \big(C^T((a_T^+,a_T^-),t-s)-(1,1)\big).\phi^T(s)ds\big).$$

Now let

$$Y^T(a,.) = \big(Y^{T,+}(a,.),Y^{T,-}(a,.)\big) = C^T\big((a_T^+,a_T^-),.T\big) : [0,1] \to \mathscr{M}^{\mathbf{1\times 2}}(\mathbb{C}).$$

Using a change of variables, we easily get that $Y^T(a,.)$ is solution of the equation

$$Y^T(a,t) = \exp\big(i(a_T^+,a_T^-) + T\int_0^t \big(Y^T(a,t-s)-(1,1)\big).\phi^T(Ts)ds\big) \tag{16}$$

and that

$$L^T(a_T^+,a_T^-,tT) = \exp\Big(\int_0^t \big(T^\alpha(Y^{T,+}(a,t-s)-1)+T^\alpha(Y^{T,-}(a,t-s)-1)\big)\big(T^{1-\alpha}\hat{\mu}(sT)\big)ds\Big). \tag{17}$$

Thanks to Remarks 1 and 2, it is easy to see that

$$\begin{aligned}
T^{1-\alpha}\hat{\mu}(sT) &= T^{1-\alpha}\mu_T + \xi T^{1-\alpha}\mu_T\big(\frac{T^\alpha}{\lambda}\int_{sT}^\infty \varphi(u)du + \lambda T^{-\alpha}\int_0^{sT}\varphi(u)du\big)\\
&= \mu\big(1+\frac{\xi}{\lambda}s^{-\alpha}(sT)^\alpha\int_{sT}^\infty \varphi(u)du\big) + \mu\xi\lambda T^{-\alpha}\int_0^{sT}\varphi(u)du\\
&\underset{T\to\infty}{\longrightarrow} \mu\big(1+\frac{\xi}{\lambda\Gamma(1-\alpha)}s^{-\alpha}\big).
\end{aligned}$$

Then the convergence of $T^\alpha(Y^T(a,t)-(1,1))$ to some functions $(c(a,t),d(a,t))$ solutions of Volterra-type equations is proved in Section 6.2. It is based on a Taylor expansion from (16). This will lead us to the expression of $L_p(a,t)$.

4.2 Main result

We define the fractional integral of order $r \in (0,1]$ of a function f as

$$I^r f(t) = \frac{1}{\Gamma(r)}\int_0^t (t-s)^{r-1}f(s)ds, \tag{18}$$

whenever the integral exists, and the fractional derivative of order $r \in [0,1)$ as

$$D^r f(t) = \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s) ds, \tag{19}$$

whenever it exists. The following theorem, proved in Section 6, is the main result of the paper.

Theorem 3. *Consider the rough Heston model* (3) *with a correlation between the two Brownian motions ρ satisfying $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$. For all $t \geq 0$ and fixed $a \in \mathbb{R}$, we have*

$$L_p(a,t) = \exp\big(\theta\lambda I^1 h(a,t) + V_0 I^{1-\alpha} h(a,t)\big), \tag{20}$$

where $h(a,.)$ is solution of the fractional Riccati equation

$$D^\alpha h(a,t) = \frac{1}{2}(-a^2 - ia) + \lambda(ia\rho\nu - 1)h(a,s) + \frac{(\lambda\nu)^2}{2} h^2(a,s), \quad I^{1-\alpha} h(a,0) = 0, \tag{21}$$

which admits a unique continuous solution.

Thus we have been able to obtain a semi-closed formula for the characteristic function in rough Heston models. This means that pricing of European options becomes an easy task in this model, see Section 5. For $\alpha = 1$, we retrieve the classical Heston formula. For $\alpha < 1$, the formula is almost the same. The difference is essentially only in that in the Riccati equation, the classical derivative is replaced by a fractional derivative. The drawback is that such fractional Riccati equations do not have explicit solutions. However, they can be solved numerically almost instantaneously, see Section 5. Finally, note that this strong link between Hawkes processes and (rough) Heston models is probably natural because both of them exhibit some kind of affine structure (although infinite-dimensional).

5 Numerical application

5.1 Numerical scheme

We explain in this section how to compute numerically the characteristic function of the log-price in a rough Heston model. By Theorem 3, $L_p(a,t)$ is entirely defined through the fractional Riccati equation (21)

$$D^\alpha h(a,t) = F\big(a, h(a,t)\big), \quad I^{1-\alpha} h(a,0) = 0,$$

where

$$F(a,x) = \frac{1}{2}(-a^2 - ia) + \lambda(ia\rho\nu - 1)x + \frac{(\lambda\nu)^2}{2} x^2.$$

Several schemes for solving (21) numerically can be found in the literature. Most of them are based on the idea that (21) implies the following Volterra equation

$$h(a,t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} F\big(a, h(a,s)\big) ds. \tag{22}$$

Then one develops numerical schemes for (22). Here we choose the well-known fractional Adams method investigated in [DFF02, DFF04, DF99]. The idea goes as follows. Let us write $g(a,t) = F\big(a, h(a,t)\big)$. Over a regular discrete time-grid $(t_k)_{k\in\mathbb{N}}$ with mesh Δ ($t_k = k\Delta$), we estimate

$$h(a,t_{k+1}) = \frac{1}{\Gamma(\alpha)}\int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1} g(a,s)ds$$

by

$$\frac{1}{\Gamma(\alpha)}\int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1} \hat{g}(a,s)ds,$$

where

$$\hat{g}(a,t) = \frac{t_{j+1}-t}{t_{j+1}-t_j}\hat{g}(a,t_j) + \frac{t-t_j}{t_{j+1}-t_j}\hat{g}(a,t_{j+1}), \quad t\in[t_j,t_{j+1}), \quad 0\le j\le k.$$

This corresponds to a trapezoidal discretization of the fractional integral and leads to the following scheme

$$\hat{h}(a,t_{k+1}) = \sum_{0\le j\le k} a_{j,k+1} F\big(a,\hat{h}(a,t_j)\big) + a_{k+1,k+1}F\big(a,\hat{h}(a,t_{k+1})\big), \tag{23}$$

with

$$a_{0,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)}\big(k^{\alpha+1} - (k-\alpha)(k+1)^\alpha\big),$$

$$a_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)}\big((k-j+2)^{\alpha+1} + (k-j)^{\alpha+1} - 2(k-j+1)^{\alpha+1}\big), \quad 1\le j\le k, \tag{24}$$

and

$$a_{k+1,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)}.$$

However, $\hat{h}(a,t_{k+1})$ being on both sides of (23), this scheme is implicit. Thus, in a first step, we compute a pre-estimation of $\hat{h}(a,t_{k+1})$ based on a Riemann sum that we then plug into the trapezoidal quadrature. This pre-estimation, called predictor and that we denote by $\hat{h}^P(a,t_{k+1})$ is defined by

$$\hat{h}^P(a,t_{k+1}) = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\tilde{g}(a,s)ds,$$

with

$$\tilde{g}(a,t) = \hat{g}(a,t_j), \quad t\in[t_j,t_{j+1}), \quad 0\le j\le k.$$

Therefore,

$$\hat{h}^P(a,t_{k+1}) = \sum_{0\le j\le k} b_{j,k+1}F\big(a,\hat{h}(a,t_j)\big),$$

where

$$b_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+1)}\big((k-j+1)^\alpha - (k-j)^\alpha\big), \quad 0\le j\le k.$$

Thus, the final explicit numerical scheme is given by

$$\hat{h}(a,t_{k+1}) = \sum_{0\le j\le k} a_{j,k+1}F\big(a,\hat{h}(a,t_j)\big) + a_{k+1,k+1}F\big(a,\hat{h}^P(a,t_j)\big), \quad \hat{h}(a,0)=0,$$

where the weights $a_{j,k+1}$ are defined in (24). Theoretical guarantees for the convergence of this scheme are provided in [LT09]. In particular, it is shown that for given $t > 0$ and $a \in \mathbb{R}$,

$$\max_{t_j \in [0,t]} |\hat{h}(a, t_j) - h(a, t_j)| = o(\Delta)$$

and

$$\max_{t_j \in [\varepsilon,t]} |\hat{h}(a, t_j) - h(a, t_j)| = o(\Delta^{2-\alpha}),$$

for any $\varepsilon > 0$.

5.2 Numerical illustrations

To compute $L_p(a, t)$, we use the numerical scheme presented above to solve fractional Riccati equations and then plug the numerical solutions into (20). Once the characteristic function is obtained, classical methods are available to obtain call prices

$$C(K, T) = \mathbb{E}[(S_T - K)_+],$$

see [CM99, Itk05, Lew01] and the survey [Sch10]. In our case, we use Lewis method, see [Lew01].

The most costly operation is the computation of $(\hat{h}(a_l, t_j), 1 \leq l \leq N_a, 1 \leq j \leq n)$ from the scheme of Section 5.1, where $n = T/\Delta$ is the number of time steps and N_a is the number of space steps a_l used for the Fourier type method finally leading to $C(T, K)$. Hence the complexity of call price computation is $O(N_a n^2)$. Note that for the classical Heston model ($\alpha = 1$), $h(a, t)$ has an explicit form and the complexity is then reduced to $O(N_a)$.

We now give a calibration result on the S&P implied volatility surface of 7 January 2010. We obtain the following optimal parameters for the rough Heston model (3):

$$\alpha = 0.62, \quad \lambda = 0.1, \quad \rho = -0.681, \quad V_0 = 0.0392, \quad \nu = 0.331, \quad \theta = 0.3156.$$

We compare model and market implied volatility surfaces in Figure II.1.

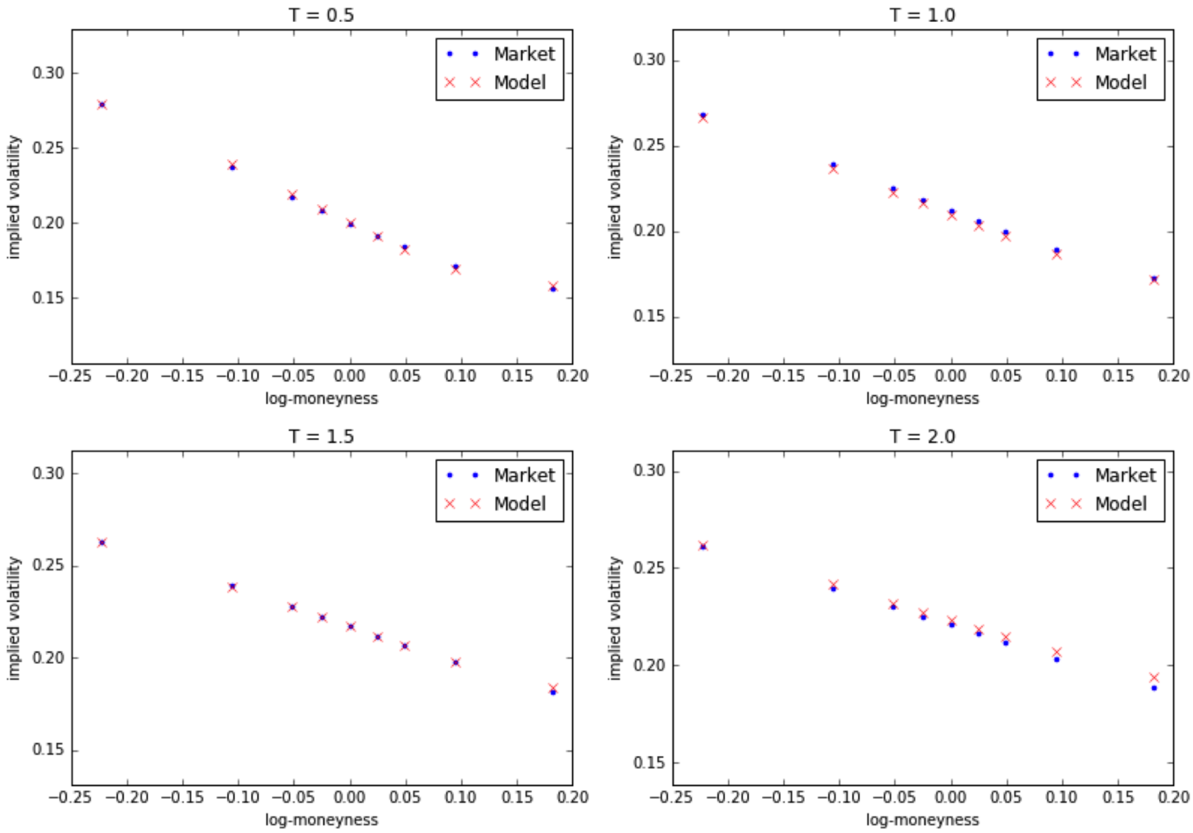

Figure II.1 – Implied volatility surface calibration with a rough-Heston model

We see that the rough Heston model provides remarkable fits for the smile, for all the considered maturities. Again, one very important point here is that the model volatility surface can be computed very efficiently thanks to our procedure.

Finally, we display the term structure of the at-the-money skew, that is the derivative of the implied volatility with respect to log-strike for at-the-money calls. We compute it for $\alpha = 1$ (classical Heston) and $\alpha = 0.62$ (rough Heston with optimal Hurst parameter equal to 0.12).

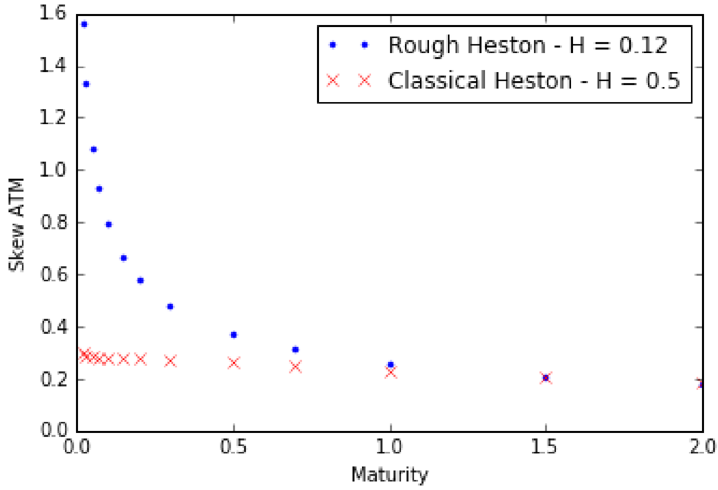

Figure II.2 – At-the-money skew as a function of maturity for $\alpha = 1$ and $\alpha = 0.62$

In the rough case, the skew explodes when maturity goes to zero, whereas it remains flat with the classical Heston model. This is a remarkable feature of rough volatility models because this exploding behavior is commonly observed on real data and very important for practical applications, see [BFG16, Fuk11, JR16b].

6 Proofs

In the sequel, c denotes a constant that may vary from line to line.

6.1 Proof of Theorem 1

The proof of Theorem 1 is close to the one given in Chapter I for the convergence of a microscopic price model to a Heston-like dynamic. The main difference is that we have to deal here with a time-varying baseline intensity $\hat{\mu}_T$, which we have introduced to get a non-zero initial volatility in the limit. As in Chapter I, we start by showing the C-tightness of (Λ^T, X^T, Z^T).

6.1.1 C-tightness of (Λ^T, X^T, Z^T)

We have the following proposition.

Proposition 1. *Under Assumptions 1 and 2, the sequence* (Λ^T, X^T, Z^T) *is C-tight and*

$$\sup_{t\in[0,1]} \|\Lambda_t^T - X_t^T\| \underset{T\to\infty}{\longrightarrow} 0$$

in probability. Moreover, if (X, Z) *is a possible limit point of* (X^T, Z^T)*, then* Z *is a continuous martingale with* $[Z, Z] = diag(X)$.

Proof:

C-tightness of X^T and Λ^T Recall that as in (9), we can write

$$\lambda_t^{T,+} = \lambda_t^{T,-} = \hat{\mu}_T(t) + \int_0^t \psi^T(t-s)\hat{\mu}_T(s)ds + \frac{1}{\beta+1}\int_0^t \psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}),$$

where

$$M_t^T = (M_t^{T,+}, M_t^{T,-}) = N_t^T - \int_0^t \lambda_s^T ds$$

is a martingale. Using that $\int_0^\cdot (f*g) = (\int_0^\cdot f)*g$, we get

$$\mathbb{E}[N_T^{T,+}] = \mathbb{E}[N_T^{T,-}] = \mathbb{E}\big[\int_0^T \lambda_s^{T,+} ds\big] = \int_0^T \hat{\mu}_T(s)ds + \int_0^T \psi^T(T-s)\big(\int_0^s \hat{\mu}_T(u)du\big)ds.$$

Consequently, $\hat{\mu}$ being a positive function and using that

$$1 + \int_0^\infty \psi^T(s)ds = 1 + \sum_{k\geq 1}\int_0^\infty (\varphi^T)^{*k} = \sum_{k\geq 0}(a_T)^k = \frac{T^\alpha}{\lambda},$$

we obtain

$$\mathbb{E}[N_T^{T,+}] \leq \int_0^T \hat{\mu}_T(s)ds\big(1 + \int_0^\infty \psi^T(s)ds\big) \leq \frac{1}{\lambda}T^{\alpha+1}\int_0^1 \hat{\mu}_T(Ts)ds.$$

Moreover, from the definition of $\hat{\mu_T}$ and Remark 1, we have

$$\int_0^1 \hat{\mu}_T(Ts)ds = \mu T^{\alpha-1}\big(1 + \xi\int_0^1 s^{-\alpha}\frac{(sT)^\alpha}{\lambda}\int_{sT}^\infty \varphi(u)duds + \lambda T^{-\alpha}\int_0^1\int_0^{sT}\varphi(u)duds\big) \leq cT^{\alpha-1}.$$

Hence $\mathbb{E}[N_T^{T,+}] \leq cT^{2\alpha}$ and therefore

$$\mathbb{E}[X_1^T] = \mathbb{E}[\Lambda_1^T] \leq c,$$

for each component. Each component of X^T and Λ^T being increasing, we deduce the tightness of each component of (X^T, Λ^T). Furthermore, the maximum jump size of X^T and Λ^T being $\frac{1-a_T}{T^\alpha \mu}$ which goes to zero, the C-tightness of (X^T, Λ^T) is obtained from Prop.VI-3.26 in [JS13].

C-tightness of Z^T It is easy to check that

$$\langle Z^T, Z^T\rangle = diag(\Lambda^T),$$

which is C-tight. From Theorem VI-4.13 in [JS13], this gives the tightness of Z^T. The maximum jump size of Z^T vanishing as T goes to infinity, we obtain that Z^T is C-tight.

Convergence of $X^T - \Lambda^T$ We have

$$X_t^T - \Lambda_t^T = \frac{1 - a_T}{T^\alpha \mu} M_{tT}^T.$$

From Doob's inequality, we get that for each component

$$\mathbb{E}[\sup_{t\in[0,1]} |\Lambda_t^T - X_t^T|^2] \leq cT^{-4\alpha}\mathbb{E}[M_T^T]^2.$$

Because $[M^T, M^T] = N^T$, we deduce

$$\mathbb{E}[\sup_{t\in[0,1]} |\Lambda_t^T - X_t^T|^2] \leq cT^{-4\alpha}\mathbb{E}[N_T^T] \leq cT^{-2\alpha}.$$

This gives the uniform convergence to zero in probability of $X^T - \Lambda^T$.

Limit of Z^T Let (X, Z) be a limit point of (X^T, Z^T). We know that (X, Z) is continuous and from Corollary IX-1.19 in [JS13], Z is a local martingale. Moreover, because

$$[Z^T, Z^T] = diag(X^T),$$

using Theorem VI-6.26 in [JS13], we get that $[Z, Z]$ is the limit of $[Z^T, Z^T]$ and $[Z, Z] = diag(X)$. By Fatou's lemma, the expectation of $[Z, Z]$ is finite and therefore Z is a martingale. □

6.1.2 Convergence of X^T and Z^T

First remark that because

$$\sup_{t\in[0,1]} |\Lambda_t^T - X_t^T| \underset{T\to\infty}{\longrightarrow} 0$$

and

$$\Lambda_t^{T,+} = \Lambda_t^{T,-},$$

we get

$$\sup_{t\in[0,1]} |X_t^{T,+} - X_t^{T,-}| \underset{T\to\infty}{\longrightarrow} 0.$$

Therefore, if a subsequence of $X_t^{T,+}$ converges to some X, then the associated subsequence of $X_t^{T,-}$ converges to the same X. We have the following proposition for the limit points of $X_t^{T,+}$ and $X_t^{T,-}$.

Proposition 2. *If (X, X, Z^+, Z^-) is a possible limit point for $(X^{T,+}, X^{T,-}, Z^{T,+}, Z^{T,-})$, then (X_t, Z_t^+, Z_t^-) can be written*

$$X_t = \int_0^t Y_s ds, \quad Z_t^+ = \int_0^t \sqrt{Y_s} dB_s^1, \quad Z_t^- = \int_0^t \sqrt{Y_s} dB_s^2,$$

where (B_1, B_2) *is a bi-dimensional Brownian motion and* Y *is solution of*

$$Y_t = \xi\big(1 - F^{\alpha,\lambda}(t)\big) + F^{\alpha,\lambda}(t) + \sqrt{\frac{1+\beta^2}{\lambda\mu(1+\beta)^2}} \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s}\,dB_s, \tag{25}$$

with

$$B = \frac{B^1 + \beta B^2}{\sqrt{1+\beta^2}}.$$

Furthermore, for any $\varepsilon > 0$, Y *has Hölder regularity* $\alpha - 1/2 - \varepsilon$.

Proof:

A convenient equality We first show the following equality

$$\hat{\mu}_T(t) + \int_0^t \psi^T(t-s)\hat{\mu}_T(s)ds = \mu_T + \xi\mu_T\frac{1}{1-a_T} + \mu_T(1-\xi)\int_0^t \psi^T(t-s)ds. \tag{26}$$

To obtain this result, we consider (26) as an equation with unknown $\hat{\mu}_T$. From Lemma 1, it admits a solution. We now look for necessary condition for this solution, showing in the end that the specific $\hat{\mu}_T$ given in Assumption 2 is the only possible choice. Using convolution by φ^T and the fact that $\psi^T * \varphi^T = \psi^T - \varphi^T$, we obtain from the left-hand side of (26):

$$\begin{aligned}
&\int_0^t \hat{\mu}_T(s)\varphi^T(t-s)ds + \int_0^t\int_0^s \psi^T(s-u)\hat{\mu}_T(u)du\varphi^T(t-s)ds\\
&= \int_0^t \hat{\mu}_T(s)\varphi^T(t-s)ds + \int_0^t\int_0^{t-u} \psi^T(s)\varphi^T(t-u-s)ds\hat{\mu}_T(u)du\\
&= \int_0^t \hat{\mu}_T(s)\varphi^T(t-s)ds + \int_0^t \big(\psi^T(t-u) - \varphi^T(t-u)\big)\hat{\mu}_T(u)du\\
&= \int_0^t \psi^T(t-s)\hat{\mu}_T(s)ds.
\end{aligned}$$

From the right-hand side of (26), we get

$$\begin{aligned}
&\int_0^t \varphi^T(t-s)(\mu_T + \xi\mu_T\frac{1}{1-a_T})ds + \mu_T(1-\xi)\int_0^t \varphi^T(t-s)\int_0^s \psi^T(s-u)duds\\
&= \mu_T(1+\xi\frac{1}{1-a_T})\int_0^t \varphi^T(t-s)ds + \mu_T(1-\xi)\int_0^t\int_0^{t-u} \psi^T(s)\varphi^T(t-u-s)dsdu\\
&= \mu_T(1+\xi\frac{1}{1-a_T})\int_0^t \varphi^T(t-s)ds + \mu_T(1-\xi)\int_0^t \big(\psi^T(t-u) - \varphi^T(t-u)\big)du.
\end{aligned}$$

Consequently, we necessarily have

$$\int_0^t \psi^T(t-s)\hat{\mu}_T(s)ds = \mu_T\xi(\frac{1}{1-a_T}+1)\int_0^t \varphi^T(t-s)ds + \mu_T(1-\xi)\int_0^t \psi^T(t-s)ds.$$

This last equation together with (26) gives that the only possible choice is

$$\hat{\mu}_T(t) = \mu_T + \xi\mu_T\frac{1}{1-a_T}\big(1 - \int_0^t \varphi^T(t-s)ds\big) - \mu_T\xi\int_0^t \varphi^T(t-s)ds.$$

End of the proof of Proposition 2 Recall that $\lambda_t^{T,+} = \lambda_t^{T,-}$. Note that using similar computations as in Section 2.2 together with (26) we can write

$$\lambda_t^{T,+} = \mu_T + \mu_T \int_0^t \psi^T(t-s)ds + \xi\mu_T\big(\frac{1}{1-a_T} - \int_0^t \psi^T(t-s)ds\big) + \frac{1}{\beta+1}\int_0^t \psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}).$$

Then using Fubini theorem together with the fact that $\int_0^{\cdot}(f*g) = (\int_0^{\cdot} f)*g$, we get

$$\int_0^t \lambda_s^{T,+}ds = \mu_T t + \mu_T \int_0^t \psi^T(t-s)sds + \xi\mu_T\big(\frac{t}{1-a_T} - \int_0^t \psi^T(t-s)sds\big)$$
$$+ \frac{1}{\beta+1}\int_0^t \psi^T(t-s)(M_s^{T,+} + \beta M_s^{T,-})ds.$$

Therefore, for $t \in [0,1]$, we have the decomposition

$$\Lambda_t^{T,+} = \Lambda_t^{T,-} = T_1 + T_2 + T_3, \tag{27}$$

with

$$T_1 = (1-a_T)t,$$
$$T_2 = T(1-a_T)\int_0^t \psi^T\big(T(t-s)\big)sds + \xi\big(t - T(1-a_T)\int_0^t \psi^T\big(T(t-s)\big)sds\big),$$
$$T_3 = \frac{1}{\sqrt{\lambda\mu(1+\beta)^2}}\int_0^t T(1-a_T)\psi^T\big(T(t-s)\big)(Z_s^{T,+} + \beta Z_s^{T,-})ds.$$

Now recall that we have shown in (10) that

$$T(1-a_T)\psi(T.) = a_T f^{\alpha,\lambda}.$$

Thus

$$T_2 \underset{T\to\infty}{\longrightarrow} \int_0^t f^{\alpha,\lambda}(t-s)sds + \xi\big(t - \int_0^t f^{\alpha,\lambda}(t-s)sds\big)$$

and

$$T_3 \underset{T\to\infty}{\longrightarrow} \frac{1}{\sqrt{\lambda\mu(1+\beta)^2}}\int_0^t f^{\alpha,\lambda}(t-s)(Z_s^+ + \beta Z_s^-)ds.$$

Therefore, letting T go to infinity in (27), we obtain using Proposition 1 that X satisfies

$$X_t = \int_0^t f^{\alpha,\lambda}(t-s)sds + \xi\big(t - \int_0^t f^{\alpha,\lambda}(t-s)sds\big) + \frac{1}{\sqrt{\lambda\mu(1+\beta)^2}}\int_0^t f^{\alpha,\lambda}(t-s)(Z_s^+ + \beta Z_s^-)ds.$$

In the same way as for the proof of Theorem 3.2 in [JR16b], we show that

$$X_t = \int_0^t Y_s ds,$$

where Y satisfies

$$Y_t = F^{\alpha,\lambda}(t) + \xi\big(1 - F^{\alpha,\lambda}(t)\big) + \frac{1}{\sqrt{\lambda\mu(1+\beta)^2}}\int_0^t f^{\alpha,\lambda}(t-s)(dZ_s^+ + \beta dZ_s^-).$$

Because, by Proposition 1,

$$[Z,Z] = \int_0^t Y_s ds \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we can apply Theorem V-3.9 in [RY13] to show the existence of a bi-dimensional Brownian motion (B^1, B^2) such that

$$Z_t^+ = \int_0^t \sqrt{Y_s} dB_s^1, \quad Z_t^- = \int_0^t \sqrt{Y_s} dB_s^2.$$

Finally, we define the following Brownian motion:

$$B = \frac{B^1 + \beta B^2}{\sqrt{1+\beta^2}}.$$

Then, in the same way as for the proof of Theorem 3.2 in [JR16b], we get that Y satisfies

$$Y_t = F^{\alpha,\lambda}(t) + \xi\big(1 - F^{\alpha,\lambda}(t)\big) + \sqrt{\frac{1+\beta^2}{\lambda\mu(1+\beta)^2}} \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s} dB_s,$$

and has Hölder regularity $\alpha - 1/2 - \varepsilon$ for any $\varepsilon > 0$. □

6.1.3 End of the proof of Theorem 1

We now recall the following proposition stating that the process Y is uniquely defined by Equation (25) and that this equation is equivalent to that given in Theorem 1. The proof of this result can be found in Chapter I. Theorem 1 is readily obtained from this proposition together with Proposition 1 and 2.

Proposition 3. *Let λ, ν, θ and V_0 be positive constants, $\alpha \in (1/2, 1)$ and B be a Brownian motion. The process V is solution of the following fractional stochastic differential equation*

$$V_t = V_0\big(1 - F^{\alpha,\lambda}(t)\big) + \theta F^{\alpha,\lambda}(t) + \nu \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s} dB_s$$

if and only if it is solution of

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda(\theta - V_s) ds + \frac{\lambda\nu}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} dB_s.$$

Furthermore, both equations admit a unique weak solution.

6.1.4 Proof of Corollary 1

From Theorem 1, we know that P^T converges in law for the Skorokhod topology to the process P given by

$$P_t = \sqrt{\frac{\theta}{2}} \int_0^t \sqrt{Y_s}(dB_s^1 - dB_s^2) - \frac{\theta}{2} \int_0^t Y_s ds.$$

Let $V_t = \theta Y_t$ and $W_t = \frac{1}{\sqrt{2}}(B_t^1 - B_t^2)$. Then

$$P_t = \int_0^t \sqrt{V_s} dW_s - \frac{1}{2}\int_0^t V_s ds,$$

where

$$V_t = \xi\theta + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(\theta - V_s)ds + \lambda\sqrt{\frac{\theta(1+\beta^2)}{\lambda\mu(1+\beta)^2}}\frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{V_s}dW_s'$$

and (W, B) is a correlated bi-dimensional Brownian motion with

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

6.2 Proof of Theorem 3

We now give the proof of Theorem 3. We do it for $t \in [0,1]$ but the proof can obviously be extended for any $t \geq 0$. We start by controlling the process $Y^T(a,t) - (1,1)$. In the sequel, $c(a)$ denotes a positive constant independent of t and T that may vary from line to line.

6.2.1 Control of $Y^T(a,t) - (1,1)$

We have the following proposition.

Proposition 4. *For any* $t \in [0,1]$,

$$T^\alpha \| Y^T(a,t) - (1,1)\| \leq c(a).$$

Proof:

Let us show that

$$T^\alpha |Y^{T,+}(a,t) - 1| \leq c(a).$$

Recall that $Y^T(a,t)$ is defined in Section 4.1 for $a \in \mathbb{R}$ by

$$Y^T(a,t) = \big(Y^{T,+}(a,t), Y^{T,-}(a,t)\big) = \big(C^{T,+}((a_T^+, a_T^-), tT), C^{T,-}((a_T^+, a_T^-), tT)\big),$$

with

$$a_T^+ = a\sqrt{\frac{\lambda\theta}{2\mu}}T^{-\alpha} - a\frac{\lambda\theta}{2\mu}T^{-2\alpha}, \quad a_T^- = -a\sqrt{\frac{\lambda\theta}{2\mu}}T^{-\alpha}.$$

Using the elements in the proof of Theorem 2 in Section 3.3, we get that

$$C^{T,+}\big((a,b),t\big) = \mathbb{E}\big[\exp(ia + ia\tilde{N}_t^{T,+} + ib\tilde{N}_t^{T,-})\big],$$

where $\tilde{N}^{T,} = (\tilde{N}^{T,+}, \tilde{N}^{T,-})$ is a bi-dimensional Hawkes process with intensity $(\tilde{\lambda}^T, \tilde{\lambda}^T)$ given by

$$\tilde{\lambda}_t^T = \frac{1}{\beta+1}\varphi^T(t) + \frac{1}{\beta+1}\int_0^t \varphi^T(t-s)(d\tilde{N}_s^{T,+} + \beta d\tilde{N}_s^{T,-}).$$

As already seen, using Lemma 1, we can rewrite the intensity under the following form

$$\tilde{\lambda}_t^T = \frac{1}{\beta+1}\psi^T(t) + \frac{1}{\beta+1}\int_0^t \psi^T(t-s)(d\tilde{M}_s^{T,+} + \beta d\tilde{M}_s^{T,-}),$$

where $\tilde{M}^T = (\tilde{M}^{T,+}, \tilde{M}^{T,-}) = \tilde{N}^T - \int_0^{\cdot} \tilde{\lambda}^T(s)ds(1,1)$ is a martingale. Using Fubini theorem, we get

$$\int_0^{tT} \tilde{\lambda}_s^T ds = \frac{1}{\beta+1} T\int_0^t \psi^T(Ts)ds + \frac{1}{\beta+1}\int_0^t T\psi^T\big(T(t-s)\big)(\tilde{M}_{sT}^{T,+} + \beta\tilde{M}_{sT}^{T,-})ds.$$

Then, from (10), we derive

$$\int_0^{tT} \tilde{\lambda}_s^T ds = \frac{1}{\lambda(\beta+1)} a_T T^\alpha F^{\alpha,\lambda}(t) + \frac{1}{\lambda(\beta+1)} a_T T^\alpha \int_0^t f^{\alpha,\lambda}(t-s)(\tilde{M}_{sT}^{T,+} + \beta\tilde{M}_{sT}^{T,-})ds. \quad (28)$$

Consequently,

$$\mathbb{E}\Big[\int_0^{tT} \tilde{\lambda}_s^T ds\Big] \leq \frac{1}{\lambda(\beta+1)} F^{\alpha,\lambda}(1) T^\alpha.$$

Let us now set $\tilde{X}_t^T = a_T^+ \tilde{N}_{tT}^{T,+} + a_T^- \tilde{N}_{tT}^{T,-}$. Using the last inequality, we deduce

$$|\mathbb{E}\tilde{X}_t^T| \leq c|a| T^{-\alpha} F^{\alpha,\lambda}(1).$$

Now recall that

$$T^\alpha(Y^{T,+}(a,t) - 1) = T^\alpha\big(\mathbb{E}\big[\exp(ia_T^+ + ia_T^+\tilde{N}_{tT}^{T,+} + ia_T^-\tilde{N}_{tT}^{T,-})\big] - 1\big).$$

Using the fact that there exists $c > 0$ such that for any $x \in \mathbb{R}$,

$$|\exp(ix) - 1 - ix| \leq c|x|^2,$$

we obtain

$$\begin{aligned}
T^\alpha|Y^{T,+}(a,t) - 1| &= T^\alpha|\mathbb{E}\big[\exp(ia_T^+ + i\tilde{X}_t^T) - 1 - i\tilde{X}_t^T - ia_T^+ + i\tilde{X}_t^T + ia_T^+\big]| \\
&\leq T^\alpha|\mathbb{E}[\tilde{X}_t^T]| + T^\alpha|a_T^+| + T^\alpha\mathbb{E}\big[|\exp(ia_T^+ + i\tilde{X}_t^T) - 1 - i\tilde{X}_t^T - ia_T^+|\big] \\
&\leq c(a)\big(1 + T^\alpha(a_T^+)^2 + T^\alpha\mathbb{E}[(\tilde{X}_t^T)^2]\big) \\
&\leq c(a)\big(1 + T^\alpha\mathbb{E}[(\tilde{X}_t^T)^2]\big).
\end{aligned}$$

Then, using that

$$\tilde{X}_t^T = a\sqrt{\frac{\lambda\theta}{2\mu}} T^{-\alpha}(\tilde{N}_{tT}^{T,+} - \tilde{N}_{tT}^{T,-}) - a\frac{\lambda\theta}{2\mu} T^{-2\alpha}\tilde{N}_{tT}^{T,+}$$

together with the fact that $\tilde{N}^{T,+} - \tilde{N}^{T,-} = \tilde{M}^{T,+} - \tilde{M}^{T,-}$, we deduce

$$T^\alpha\mathbb{E}[(\tilde{X}_t^T)^2] \leq ca^2 T^{-\alpha}\mathbb{E}[(\tilde{M}_{tT}^{T,+} - \tilde{M}_{tT}^{T,-})^2] + ca^2 T^{-3\alpha}\mathbb{E}[(\tilde{N}_{tT}^{T,+})^2].$$

Because $[\tilde{M}^{T,+} - \tilde{M}^{T,-}, \tilde{M}^{T,+} - \tilde{M}^{T,-}] = \tilde{N}^{T,+} + \tilde{N}^{T,-}$, we get

$$\begin{aligned} T^\alpha \mathbb{E}[(\tilde{X}_t^T)^2] &\le ca^2 T^{-\alpha} \mathbb{E}[\tilde{N}_{tT}^{T,+} + \tilde{N}_{tT}^{T,-}] + ca^2 T^{-3\alpha} \mathbb{E}[(\tilde{N}_{tT}^{T,+})^2] \\ &\le ca^2 \big(T^{-\alpha} \mathbb{E}[\int_0^{tT} \tilde{\lambda}_s^T ds] + T^{-3\alpha} \mathbb{E}[(\tilde{N}_{tT}^{T,+})^2] \big). \\ &\le ca^2 \big(1 + T^{-3\alpha} \mathbb{E}[(\tilde{N}_{tT}^{T,+})^2] \big). \end{aligned}$$

In order to control the term $\mathbb{E}[(\tilde{N}_{tT}^{T,+})^2]$, we now compute a bound for $\mathbb{E}\big[(\int_0^{tT} \tilde{\lambda}_s^T ds)^2\big]$. Using (28), this last quantity is equal to

$$\frac{1}{\lambda^2(\beta+1)^2} a_T^2 T^{2\alpha} \big(F^{\alpha,\lambda}(t)\big)^2 + \frac{1}{\lambda^2(\beta+1)^2} a_T^2 T^{2\alpha} \mathbb{E}\Big[\Big(\int_0^t f^{\alpha,\lambda}(t-s)(\tilde{M}_{sT}^{T,+} + \beta \tilde{M}_{sT}^{T,-}) ds\Big)^2\Big],$$

which is smaller than

$$c(a) T^{2\alpha} \Big(1 + \mathbb{E}\big[\int_0^t \big(f^{\alpha,\lambda}(t-s)\big)^2 (\tilde{M}_{sT}^{T,+} + \beta \tilde{M}_{sT}^{T,-})^2 ds\big]\Big).$$

Because $[\tilde{M}^{T,+} + \beta \tilde{M}^{T,-}, \tilde{M}^{T,+} + \beta \tilde{M}^{T,-}] = \tilde{N}^{T,+} + \beta^2 \tilde{N}^{T,-}$, we obtain

$$\begin{aligned} \mathbb{E}\big[(\int_0^{tT} \tilde{\lambda}_s^T ds)^2\big] &\le c(a) T^{2\alpha} \Big(1 + \int_0^t \big(f^{\alpha,\lambda}(t-s)\big)^2 \mathbb{E}[\tilde{N}_{sT}^{T,+} + \beta^2 \tilde{N}_{sT}^{T,-}] ds \Big) \\ &\le c(a) T^{2\alpha} \Big(1 + \int_0^t \big(f^{\alpha,\lambda}(t-s)\big)^2 \mathbb{E}\big[\int_0^{sT} \tilde{\lambda}_u^T du\big] ds \Big) \\ &\le c(a) T^{2\alpha} \Big(1 + T^\alpha \int_0^1 \big(f^{\alpha,\lambda}(s)\big)^2 ds \Big) \\ &\le c(a) T^{3\alpha}. \end{aligned}$$

Thus

$$\mathbb{E}\big[(\tilde{N}_{tT}^{T,+})^2\big] \le 2\mathbb{E}\big[(\tilde{M}_{tT}^{T,+})^2\big] + 2\mathbb{E}\big[\big(\int_0^{tT} \tilde{\lambda}_s^T ds\big)^2\big] \le c(a) T^{3\alpha}.$$

Finally, $T^\alpha \mathbb{E}[(\tilde{X}_t^T)^2] \le c(a)$ and therefore

$$T^\alpha |Y^{T,+}(a,t) - 1| \le c(a).$$

The fact that

$$T^\alpha |Y^{T,-}(a,t) - 1| \le c(a)$$

is proved similarly. □

6.2.2 Convergence of $T^\alpha(Y^T - (1,1))$

Let $\kappa = \lambda\theta/(2\mu)$. We have the following proposition.

Proposition 5. *The sequence $T^\alpha(Y^T(a,t)-(1,1))$ converges uniformly in $t\in[0,1]$ to $\big(c(a,t),d(a,t)\big)$, where (c,d) are solutions of*

$$c(a,t) = ia\sqrt{\kappa} - ia\frac{\kappa}{\lambda(\beta+1)}F^{\alpha,\lambda}(t) + \frac{1}{2\lambda(\beta+1)}\int_0^t \big(c^2(a,t-s)+d^2(a,t-s)\big)f^{\alpha,\lambda}(s)ds$$

$$d(a,t) = -ia\sqrt{\kappa} - ia\frac{\beta\kappa}{\lambda(\beta+1)}F^{\alpha,\lambda}(t) + \frac{\beta}{2\lambda(\beta+1)}\int_0^t \big(c^2(a,t-s)+d^2(a,t-s)\big)f^{\alpha,\lambda}(s)ds.$$

Proof:

Convenient rewriting of $T^\alpha(Y^T-(1,1))$ Using the fact that the complex logarithm[5] is analytic on the set $\mathbb{C}/\mathbb{R}^-$, we can show that there exists $c>0$ such that for any $x\in\mathbb{C}$ with $|x|<1/2$,

$$|\log(1+x)-x+\frac{1}{2}x^2|\leq c|x|^3.$$

Thus we can write

$$\log\big(Y^T(a,t)\big) = Y^T(a,t)-(1,1)-\frac{1}{2}\big(Y^T(a,t)-(1,1)\big)^2-\varepsilon^T(a,t),$$

with $|\varepsilon^T(a,t)|\leq c(a)T^{-3\alpha}$. Indeed, for large enough T, we have from Proposition 4 that $|Y^{T,+}(a,t)-1|\leq 1/2$ and $|Y^{T,-}(a,t)-1|\leq 1/2$, uniformly in t. Now, again from Proposition 4, it is easy to see that

$$\|i(a_T^+,a_T^-)+\int_0^t T\big(Y^T(a,t-s)-(1,1)\big).\phi^T(Ts)ds\|\leq c(a)T^{-\alpha}\underset{T\to\infty}{\longrightarrow}0.$$

Hence, for large enough T, the imaginary part of

$$i(a_T^+,a_T^-)+\int_0^t T(Y^T(a,t-s)-(1,1)).\phi^T(Ts)ds$$

has a norm which is smaller than π. Therefore

$$\log\Big(\exp\big(i(a_T^+,a_T^-)+\int_0^t T\big(Y^T(a,t-s)-(1,1)\big).\phi^T(Ts)ds\big)\Big)$$

is equal to

$$i(a_T^+,a_T^-)+\int_0^t T\big(Y^T(a,t-s)-(1,1)\big).\phi^T(Ts)ds.$$

Then, using Equation (16), we get

$$\begin{aligned}Y^T(a,t)-(1,1) &= \frac{1}{2}\big(Y^T(a,t)-(1,1)\big)^2+\varepsilon^T(a,t)+ia\sqrt{\kappa}T^{-\alpha}(1,-1)\\ &\quad -ia\kappa T^{-2\alpha}(1,0)+T\int_0^t\big(Y^T(a,t-s)-(1,1)\big).\phi^T(Ts)ds.\end{aligned}$$

[5] The complex logarithm is defined on $\mathbb{C}/\mathbb{R}^-$ by $\log(z)=\log(|z|)+i\arg(z)$, with $\arg(z)\in(-\pi,\pi]$.

Using again the fact that

$$\sum_{k\geq 1} \big(T\phi^T(T.)\big)^{*k} = a_T \frac{T^\alpha}{\lambda} f^{\alpha,\lambda}\chi,$$

together with Lemma 1, we derive

$$\begin{aligned} Y^T(a,t)-(1,1) &= \frac{1}{2}\big(Y^T(a,t)-(1,1)\big)^2 + \varepsilon^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) - ia\kappa T^{-2\alpha}(1,0) \\ &+ \frac{a_T}{2}\frac{T^\alpha}{\lambda}\int_0^t \big(Y^T(a,t-s)-(1,1)\big)^2.\chi f^{\alpha,\lambda}(s)ds + \frac{a_T}{\lambda}T^\alpha \int_0^t \varepsilon^T(a,t-s).\chi f^{\alpha,\lambda}(s)ds \\ &+ ia\sqrt{\kappa}\frac{a_T}{\lambda}(1,-1).\chi F^{\alpha,\lambda}(t) - ia\kappa T^{-\alpha}\frac{a_T}{\lambda}(1,0).\chi F^{\alpha,\lambda}(t). \end{aligned}$$

Let

$$\varepsilon_1^T(a,t) = \frac{1}{2}\big(Y^T(a,t)-(1,1)\big)^2 + \varepsilon^T(a,t) - ia\kappa T^{-2\alpha}(1,0) + \frac{a_T}{\lambda}T^\alpha\int_0^t \varepsilon^T(a,t-s).\chi f^{\alpha,\lambda}(s)ds.$$

We have

$$\begin{aligned} Y^T(a,t)-(1,1) &= \varepsilon_1^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) + \frac{a_T}{2}\frac{T^\alpha}{\lambda}\int_0^t \big(Y^T(a,t-s)-(1,1)\big)^2.\chi f^{\alpha,\lambda}(s)ds \\ &- ia_T a\frac{\kappa}{\lambda(\beta+1)}T^{-\alpha}F^{\alpha,\lambda}(t)(1,\beta). \end{aligned}$$

Let now

$$\varepsilon_2^T(a,t) = -\frac{1}{2}\int_0^t \big(Y^T(a,t-s)-(1,1)\big)^2.\chi f^{\alpha,\lambda}(s)ds + ia\frac{\kappa}{(\beta+1)}T^{-2\alpha}F^{\alpha,\lambda}(t)(1,\beta).$$

We obtain

$$\begin{aligned} Y^T(a,t)-(1,1) &= \varepsilon_1^T(a,t) + \varepsilon_2^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) + \frac{1}{2\lambda}T^\alpha\int_0^t \big(Y^T(a,t-s)-(1,1)\big)^2.\chi f^{\alpha,\lambda}(s)ds \\ &- ia\frac{\kappa}{\lambda(\beta+1)}T^{-\alpha}F^{\alpha,\lambda}(t)(1,\beta). \end{aligned}$$

Using Proposition 4, we easily see that $T^{2\alpha}\varepsilon_1^T$ and $T^{2\alpha}\varepsilon_2^T$ are uniformly bounded in t and T. We now set

$$\theta^T(a,t) = \big(\theta^{T,+}(a,t), \theta^{T,-}(a,t)\big) = T^\alpha\big(Y^T(a,t)-(1,1)\big)$$

and

$$r^T(a,t) = T^\alpha\big(\varepsilon_1^T(a,t) + \varepsilon_2^T(a,t)\big).$$

We have that $T^\alpha r^T$ is uniformly bounded in t and T and

$$\theta^T(a,t) = r^T(a,t) + ia\sqrt{\kappa}(1,-1) - ia\frac{\kappa}{\lambda(\beta+1)}F^{\alpha,\lambda}(t)(1,\beta) + \frac{1}{2\lambda}\int_0^t \big(\theta^T(a,t-s)\big)^2.\chi f^{\alpha,\lambda}(s)ds.$$

Convergence of θ^T For fixed a, we now show that $t \to \theta^T(a,t)$ is a Cauchy sequence in the space of continuous functions $C([0,1],\mathbb{R}^2)$ equipped with the sup-norm. Let $\delta > 0$ and $T_0 > 1$ such that for $T > T_0$, $\|r^T(a,t)\|_\infty \le \frac{\delta}{2}$ for any $t \in [0,1]$. Then for $T > T_0$, $T' > T_0$ and $t \in [0,1]$,

$$\|\theta^T(a,t) - \theta^{T'}(a,t)\| \le \delta + \frac{1}{2\lambda}\int_0^t \|\big(\theta^T(a,t-s)\big)^2.\chi - \big(\theta^{T'}(a,t-s)\big)^2.\chi\| f^{\alpha,\lambda}(s)ds.$$

Because θ^T is uniformly bounded in t and T, we get

$$\|\theta^T(a,t) - \theta^{T'}(a,t)\| \le \delta + C(a)\int_0^t \|\theta^T(a,t-s) - \theta^{T'}(a,t-s)\| f^{\alpha,\lambda}(s)ds.$$

Using Lemma 3 in Appendix, this enables us to show that θ^T is a Cauchy sequence. Consequently, $\theta^T(a,t)$ converges uniformly in t to $\big(c(a,t), d(a,t)\big)$, where (c,d) is solution to the following equation

$$c(a,t) = ia\sqrt{\kappa} - ia\frac{\kappa}{\lambda(\beta+1)}F^{\alpha,\lambda}(t) + \frac{1}{2\lambda(\beta+1)}\int_0^t \big(c^2(a,t-s) + d^2(a,t-s)\big)f^{\alpha,\lambda}(s)ds$$

$$d(a,t) = -ia\sqrt{\kappa} - ia\frac{\beta\kappa}{\lambda(\beta+1)}F^{\alpha,\lambda}(t) + \frac{\beta}{2\lambda(\beta+1)}\int_0^t \big(c^2(a,t-s) + d^2(a,t-s)\big)f^{\alpha,\lambda}(s)ds.$$

□

6.2.3 End of the proof of Theorem 3

Deriving the characteristic function Let $a \in \mathbb{R}$. Recall that from Section 4.1, we have

$$L^T(a_T^+, a_T^-, tT) = \exp\Big(\int_0^t \big(T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1)\big)\big(T^{1-\alpha}\hat{\mu}(sT)\big)ds\Big)$$

and furthermore, from Proposition 5,

$$T^\alpha(Y^{T,+}(a,t)-1) + T^\alpha(Y^{T,-}(a,t)-1)$$

converges uniformly in t to $c(a,t) + d(a,t)$. Also, using Remark 2, we have

$$T^{1-\alpha}\hat{\mu}(tT) = \mu + \mu\xi\big(\frac{t^{-\alpha}}{\lambda}(Tt)^\alpha \int_{tT}^\infty \varphi(s)ds + \lambda T^{-\alpha}\int_0^{tT}\varphi(s)ds\big)$$

and therefore $T^{1-\alpha}\hat{\mu}(tT)$ converges towards

$$\mu\big(1 + \xi\frac{t^{-\alpha}}{\lambda\Gamma(1-\alpha)}\big).$$

In addition, using Proposition 4, we get that for given $t \in [0,1]$ and for any $s \in [0,t]$

$$|T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1)|\big(T^{1-\alpha}\hat{\mu}(sT)\big) \le c(a)(1+s^{-\alpha}).$$

The right hand side of the last inequality is integrable over $[0,t]$. Therefore, using the convergence of $L^T(a_T^+, a_T^-, tT)$ towards $L_p(a,t)$ and applying the dominated convergence theorem, we obtain

$$L_p(a,t) = \exp\Big(\int_0^t g(a,s)(1+\xi\frac{(t-s)^{-\alpha}}{\lambda\Gamma(1-\alpha)})ds\Big),$$

where $g(a,t) = \mu\big(c(a,t)+d(a,t)\big)$. Thus, we have shown that

$$L_p(a,t) = \exp\Big(\int_0^t g(a,s)ds + \frac{V_0}{\theta\lambda}I^{1-\alpha}g(a,t)\Big).$$

Integral equation for g We now prove that g is solution of an integral equation. First remark that

$$d(a,t) = \beta c(a,t) - ia(1+\beta)\sqrt{\kappa}.$$

Hence $g(a,t) = \mu(\beta+1)(c(a,t)-ia\sqrt{\kappa})$, which can be written

$$-ia\frac{\mu\kappa}{\lambda}F^{\alpha,\lambda}(t) + \frac{\mu}{2\lambda}\int_0^t \Big((c(a,s)-ia\sqrt{\kappa}+ia\sqrt{\kappa})^2 + \big(\beta(c(a,s)-ia\sqrt{\kappa})-ia\sqrt{\kappa}\big)^2\Big)f^{\alpha,\lambda}(t-s)ds.$$

Thus,

$$\begin{aligned}g(a,t) = &-ia\frac{\mu\kappa}{\lambda}F^{\alpha,\lambda}(t) + \frac{1+\beta^2}{2\mu\lambda(1+\beta)^2}\int_0^t \big(g(a,s)\big)^2 f^{\alpha,\lambda}(t-s)ds - a^2\frac{\mu\kappa}{\lambda}F^{\alpha,\lambda}(t)\\ &+ia\frac{\sqrt{\kappa}(1-\beta)}{\lambda(\beta+1)}\int_0^t g(a,s)f^{\alpha,\lambda}(t-s)ds.\end{aligned}$$

Using the definition of κ in Section 6.2, we deduce

$$\begin{aligned}g(a,t) = &\frac{\theta}{2}(-a^2-ia)F^{\alpha,\lambda}(t) + ia\frac{\sqrt{\theta}(1-\beta)}{\sqrt{2\lambda\mu}(\beta+1)}\int_0^t g(a,s)f^{\alpha,\lambda}(t-s)ds\\ &+\frac{1+\beta^2}{2\mu\lambda(1+\beta)^2}\int_0^t g^2(a,s)f^{\alpha,\lambda}(t-s)ds\end{aligned}$$

and from those of ρ and ν in Section 4.1, we finally obtain that $g(a,t)$ is equal to

$$\frac{\theta}{2}(-a^2-ia)F^{\alpha,\lambda}(t) + ia\rho\nu\int_0^t g(a,s)f^{\alpha,\lambda}(t-s)ds + \frac{\nu^2}{2\theta}\int_0^t \big(g(a,s)\big)^2 f^{\alpha,\lambda}(t-s)ds.$$

Thus,

$$L_p(a,t) = \exp\Big(\int_0^t g(a,s)\big(1+\xi\frac{(t-s)^{-\alpha}}{\lambda\Gamma(1-\alpha)}\big)ds\Big)$$

with

$$g(a,t) = \int_0^t \Big(\frac{\theta}{2}(-a^2-ia) + ia\rho\nu g(a,s) + \frac{\nu^2}{2\theta}\big(g(a,s)\big)^2\Big)f^{\alpha,\lambda}(t-s)ds.$$

Let us now set $h = g/(\theta\lambda)$. Then

$$L_p(a,t) = \exp\Big(\int_0^t h(a,s)\big(\theta\lambda + V_0\frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)}\big)ds\Big),$$

with

$$h(a,t)=\int_0^t \Big(\frac{1}{2}(-a^2-ia)+ia\lambda\rho\nu h(a,s)+\frac{(\lambda\nu)^2}{2}\big(h(a,s)\big)^2\Big)\frac{1}{\lambda}f^{\alpha,\lambda}(t-s)ds. \tag{29}$$

Using Lemma 2, we have that Equation (29) can also be written under the following form

$$D^\alpha h(a,t)=\frac{1}{2}(-a^2-ia)+\lambda(ia\rho\nu-1)h(a,s)+\frac{(\lambda\nu)^2}{2}\big(h(a,s)\big)^2,\quad I^{1-\alpha}h(a,0)=0.$$

6.2.4 Uniqueness of the solution of (21)

For a given $a\in\mathbb{R}$, consider two continuous solutions $h_1(a,.)$ and $h_2(a,.)$ of (21) or equivalently of (29). We have that $|h_1(a,t)-h_2(a,t)|$ is smaller than

$$\int_0^t \big(|a\rho\nu||h_1(a,s)-h_2(a,s)|+\frac{\lambda\nu^2}{2}|\big(h_1(a,s)\big)^2-\big(h_2(a,s)\big)^2|\big)f^{\alpha,\lambda}(t-s)ds.$$

Using the continuity of $h_1(a,.)$ and $h_2(a,.)$, this is also smaller than

$$c(a)\int_0^t |h_1(a,s)-h_2(a,s)|f^{\alpha,\lambda}(t-s)ds.$$

Thanks to Lemma 3, this gives $h_1(a,.)=h_2(a,.)$.

II.A Mittag-Leffler functions

Let $(\alpha,\beta)\in(\mathbb{R}_+^*)^2$. The Mittag-Leffler function $E_{\alpha,\beta}$ is defined for $z\in\mathbb{C}$ by

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

For $(\alpha,\gamma)\in(0,1)\times\mathbb{R}_+$, we also define

$$f^{\alpha,\gamma}(t)=\gamma t^{\alpha-1}E_{\alpha,\alpha}(-\gamma t^\alpha),\quad t>0,$$

$$F^{\alpha,\gamma}=\int_0^t f^{\alpha,\gamma}(s)ds,\quad t\geq 0.$$

The function $f^{\alpha,\gamma}$ is a density function on $\mathbb{R}_+$ called Mittag-Leffler density function. The following properties of $f^{\alpha,\gamma}$ and $F^{\alpha,\gamma}$ can be found in [HMS11, Mai, MH08]. We have

$$f^{\alpha,\gamma}(t)\underset{t\to 0^+}{\sim}\frac{\gamma}{\Gamma(\alpha)}t^{\alpha-1},\quad f^{\alpha,\gamma}(t)\underset{t\to\infty}{\sim}\frac{\alpha}{\gamma\Gamma(1-\alpha)}t^{-(\alpha+1)}$$

and

$$F^{\alpha,\gamma}(t)=1-E_{\alpha,1}(-\gamma t^\alpha),\quad F^{\alpha,\gamma}(t)\underset{t\to 0^+}{\sim}\frac{\gamma}{\Gamma(\alpha+1)}t^\alpha,\quad 1-F^{\alpha,\gamma}(t)\underset{t\to\infty}{\sim}\frac{1}{\gamma\Gamma(1-\alpha)}t^{-\alpha}.$$

Finally, for $\alpha\in(1/2,1)$, $f^{\alpha,\gamma}$ is square-integrable and its Laplace transform is given for $z\geq 0$ by

$$\hat{f}^{\alpha,\gamma}(z)=\int_0^\infty f_{\alpha,\gamma}(s)e^{-zs}ds=\frac{\gamma}{\gamma+z^\alpha}.$$

II.B Wiener-Hopf equations

The following result is used extensively in this work to solve Wiener-Hopf type equations, see for example [BDHM13b].

Lemma 1. *Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^d$ and $\phi : \mathbb{R}_+ \to \mathscr{M}^d(\mathbb{R})$ be a matrix-valued function with integrable components such that $\mathscr{S}(\int_0^\infty \phi(s)ds) < 1$. Then there exists a unique locally bounded function f from $\mathbb{R}$ to $\mathbb{R}^d$ solution of*

$$f(t) = g(t) + \int_0^t \phi(t-s).f(s)ds, \quad t \geq 0$$

given by

$$f(t) = g(t) + \int_0^t \psi(t-s).g(s)ds, \quad t \geq 0,$$

where $\psi = \sum\limits_{k\geq 1} \phi^{*k}$.

II.C Fractional differential equations

We end this appendix with some useful results about fractional differential equations. The next lemma can be found in [SKM93].

Lemma 2. *Let h be a continuous function from $[0,1]$ to $\mathbb{R}$, $\alpha \in (0,1]$ and $\gamma \in \mathbb{R}$. There is a unique continuous solution to the equation*

$$D^\alpha y(t) = \gamma y(t) + h(t), \quad I^{1-\alpha} y(0) = 0$$

given by

$$y(t) = \int_0^t (t-s)^{\alpha-1} E_{\alpha,\alpha}\big(\gamma(t-s)^\alpha\big) h(s)ds.$$

We also have the following useful result.

Lemma 3. *Let h be a non-negative continuous function from $[0,1]$ to $\mathbb{R}$ such that for any $t \in [0,1]$,*

$$h(t) \leq \varepsilon + C \int_0^t f^{\alpha,\gamma}(t-s)h(s)ds,$$

for some $\varepsilon \geq 0$ and $C \geq 0$. Then for any $t \in [0,1]$,

$$h(t) \leq C'\varepsilon,$$

with

$$C' = 1 + C\gamma \int_0^1 s^{\alpha-1} E_{\alpha,\alpha}\big(\gamma(C-1)s^\alpha\big)ds > 0.$$

In particular, if $\varepsilon = 0$ then $h = 0$.

Proof:

Let
$$f(t) = h(t) - C\int_0^t f^{\alpha,\gamma}(t-s)h(s)ds.$$
and $g = h - f$. The function g is solution of
$$g(t) = C\int_0^t f^{\alpha,\gamma}(t-s)\big(g(s)+f(s)\big)ds.$$
Thus, from Lemma 2, g is the unique solution of
$$D^\alpha g(t) = \gamma(C-1)g(t) + C\gamma f(t), \quad I^{1-\alpha}g(0) = 0.$$
Hence using again Lemma 2, we deduce that
$$g(t) = C\gamma\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}\big(\gamma(C-1)(t-s)^\alpha\big)f(s)ds.$$
Therefore,
$$g(t) \le C\gamma\varepsilon\int_0^t s^{\alpha-1}E_{\alpha,\alpha}\big(\gamma(C-1)s^\alpha\big)ds.$$
Using that $h = f + g$ together with the fact that $E_{\alpha,\alpha}$ is non-negative, we get the result. □

Part III

Hedging under the rough Heston model

CHAPTER III

Perfect hedging in rough Heston models

Abstract

Rough volatility models are known to reproduce the behavior of historical volatility data while at the same time fitting the volatility surface remarkably well, with very few parameters. However, managing the risks of derivatives under rough volatility can be intricate since the dynamics involve fractional Brownian motion. We show in this paper that surprisingly enough, explicit hedging strategies can be obtained in the case of rough Heston models. The replicating portfolios contain the underlying asset and the forward variance curve, and lead to perfect hedging (at least theoretically). From a probabilistic point of view, our study enables us to disentangle the infinite-dimensional Markovian structure associated to rough volatility models.

Keywords: Rough volatility, rough Heston model, Hawkes processes, fractional Brownian motion, fractional Riccati equations, limit theorems, forward variance curve.

1 Introduction

It has been recently shown in [GJR18] that rough fractional processes enable us to reproduce very accurately the behavior of historical volatility time-series. More precisely, the dynamic of their logarithm is quite similar to that of a fractional Brownian motion with Hurst parameter of order 0.1. Recall that a fractional Brownian motion W^H with Hurst parameter $H \in (0,1)$ can be built from a classical two-sided Brownian motion W through the Mandelbrot-van Ness representation:

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

The fractional Brownian motion has Hölder regularity $H-\varepsilon$ for any $\varepsilon > 0$. Hence fractional volatility models with small Hurst parameter are referred to as rough volatility models.

Beyond historical data modeling, rough volatility models provide excellent fits and dynamics for the whole volatility surface, in particular for the at-the-money skew, with very few scalar parameters (typically three), see [BFG16, Fuk11, GJR18]. One of the only potential drawbacks of such models in practice is the difficulty to price and hedge derivatives with them. Indeed, although some promising approaches have been recently introduced, see [BLP17], due to the non-Markovian nature of the fractional Brownian motion, running efficient Monte-Carlo methods remains an intricate task in the rough volatility context, see [NS16].

However, it is shown in Chapter II that in the specific case of the so-called rough Heston model, instantaneous pricing of derivatives can be obtained. The rough Heston model of Chapter II is a natural extension[1] to the rough framework of the classical Heston model of [Hes93]. Indeed, the dynamic of the price S on a probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ is defined as follows:

$$dS_t = S_t \sqrt{V_t} dW_t$$
$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-u)^{\alpha-1} \lambda(\theta - V_u) du + \frac{1}{\Gamma(\alpha)} \int_0^t (t-u)^{\alpha-1} \nu \sqrt{V_u} dB_u. \tag{1}$$

Here the parameters λ, θ, V_0, S_0 and ν are positive, $\alpha \in (1/2, 1)$ and $W = \rho B + \sqrt{1-\rho^2} B^{\perp}$ with $(B, B^{\perp})$ a two-dimensional $\mathbb{F}$-Brownian motion and $\rho \in [-1, 1]$. From Chapter II, the fractional stochastic differential equation (1) admits a unique weak solution and this solution has sample paths with Hölder regularity $\alpha - 1/2 - \varepsilon$ almost surely, for any $\varepsilon > 0$. Note also that in the case $\alpha = 1$, we retrieve the classical Heston model. Surprisingly enough, it is proved in Chapter II that a semi-closed formula *à la Heston* also holds for the characteristic function of the log-price in the rough Heston model. This formula is very similar to that obtained in the classical Heston case, except that the classical time-derivative in the Riccati equation has to be replaced by a fractional derivative. Indeed, we have

$$\mathbb{E}[\exp\big(ia \log(S_t/S_0)\big)] = \exp\big(g_1(a,t) + V_0 g_2(a,t)\big),$$

where

$$g_1(a,t) = \theta\lambda \int_0^t h(a,s) ds, \quad g_2(a,t) = I^{1-\alpha} h(a,t),$$

and h is the unique continuous solution of the following fractional Riccati equation:

$$D^{\alpha} h(a,s) = \frac{1}{2}(-a^2 - ia) + (ia\rho\nu - \lambda) h(a,s) + \frac{\nu^2}{2} h^2(a,s), \quad I^{1-\alpha} h(a,0) = 0,$$

with $I^{1-\alpha}$ and D^{α} the fractional integral and derivative operators defined in Appendix III.A. When $\alpha = 1$, this result does coincide with the classical Heston's result. Furthermore, efficient numerical pricing procedures for vanilla options can be easily designed from it, see Chapter II.

Thus, the relevance of the rough Heston model is twofold: it enjoys at the same time the nice modeling properties of rough volatility models and the computational advantages of the Heston framework. However, the interest of having a pricing procedure is of course limited if it does not go along with a hedging strategy. Being able to build a hedging portfolio

[1] Actually there is no really standard definition for the rough Heston model and other versions can be considered, see [GJR14].

essentially means computing conditional expectations of the form $C_t = \mathbb{E}[f(S_T)|\mathscr{F}_t]$, where f is a deterministic payoff function. In the classical Heston case, the Markovian structure of the model is very helpful to do it. In the rough case, this task is much more intricate since the underlying fractional Brownian motion is neither a Markov process nor a semi-martingale.

To tackle this issue, we first study the conditional laws in rough Heston models. We actually prove a very nice stability property. Indeed, we show that conditional on $\mathscr{F}_t$, the law of the rough Heston model is still that of a rough Heston model, provided that the mean-reversion level θ is replaced by a time-dependent one. Hence we generalize our definition of the rough Heston model, allowing for the mean-reversion level to depend on time. Then using Hawkes processes as in Chapter II, we are able to compute the extended characteristic function of the log-price in generalized rough Heston models, that is

$$\mathbb{E}\big[\exp\big(z\log(S_t/S_0)\big)\big] \tag{2}$$

for $z = a + ib$, with $b \in \mathbb{R}$ and a in some subset of $\mathbb{R}$ to be defined later. From an explicit expression of (2), we can deduce a semi-closed formula for C_t, following for example the approach in [CM99].

Our most important result is the fact that we are able to identify the relevant state variables in rough Heston models, namely the underlying and the so-called forward variance curve: $(\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{0\leq s\leq T-t}$. Indeed, we show that C_t can be written

$$C_t = C\big(T-t, S_t, (\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}\big),$$

with $C()$ an explicit deterministic function. The above formula shows rigorously that the hedging instruments needed with rough models are the spot price and the forward variance curve, an idea already emphasized in [BFG16]. Such result is also in the spirit of the approach developed in [Ber05]. More precisely, we show that the dynamic of the option price satisfies

$$\begin{aligned} dC_t = {} & \partial_S C\big(T-t, S_t, (\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}\big) dS_t \\ & + \partial_V C\big(T-t, S_t, (\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}\big).\big(d\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\geq 0}\big), \end{aligned}$$

where $\partial_S C$ is the derivative of C with respect to the underlying (the so-called delta) and $\partial_V C$ is the Fréchet derivative of C according to the forward variance curve. From this expression, we readily obtain hedging strategies in terms of underlying and forward variance curve. Of course, in practice, one cannot really trade the whole forward variance curve. However, approximations can be built using liquid variance swaps or vanilla options.

Note also that using generalized rough Heston models enables us to perfectly fit the initial forward variance curve through the time varying mean-reversion parameter. Thus, one reproduces with great accuracy the dynamics of historical data, the whole implied volatility surface, including the at-the-money skew and the forward variance curve, and has access to instantaneous pricing and hedging methods.

The paper is organized as follows. In Section 2, we investigate conditional laws of rough Heston models and introduce generalized rough Heston models with time-dependent mean-reversion level. Using Hawkes processes, we derive in Section 3 the characteristic function of the log-price in generalized rough Heston models, emphasizing the role of the forward variance curve. We also discuss useful sufficient conditions for finite moments of the underlying price. Finally, we design our hedging strategies in Section 4. Some proofs are relegated to Section 5 and some technical results are given in an Appendix.

2 Conditional laws of rough Heston models

The goal of this paper is to understand how to price and hedge vanilla options with maturity $T > 0$ and payoff $f(S_T)$ in the rough Heston framework (1). Thus a first step is to characterize the law of the process $(S_t^{t_0}, V_t^{t_0})_{t\geq 0} = (S_{t+t_0}, V_{t+t_0})_{t\geq 0}$ conditional on $\mathscr{F}_{t_0}$, for a fixed $t_0 > 0$. Indeed, in order to derive the option price dynamic and to build hedging portfolios, one needs to be able to compute $\mathbb{E}[f(S_T)|\mathscr{F}_t], \quad 0 \leq t \leq T$.

To state our result on conditional laws of rough Heston models, it is convenient to introduce a generalized version of Model (1), allowing for time-varying mean-reversion level.

Definition 1 (Generalized rough Heston model). *On a filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$*, we define a generalized rough Heston model by*

$$dS_t = S_t\sqrt{V_t}dW_t$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\lambda(\theta^0(u) - V_u)du + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\nu\sqrt{V_u}dB_u. \tag{3}$$

Here the parameters λ*,* V_0*,* S_0 *and* ν *are positive,* $\alpha \in (1/2, 1)$ *and* $W = \rho B + \sqrt{1-\rho^2}B^\perp$ *with* $(B, B^\perp)$ *a two-dimensional* $\mathbb{F}$*-Brownian motion and* $\rho \in [-1, 1]$*. Moreover,* θ^0 *is a deterministic function, continuous on* $\mathbb{R}_+^*$ *satisfying*

$$\forall u > 0; \quad \theta^0(u) \geq -\frac{V_0}{\lambda\Gamma(1-\alpha)}u^{-\alpha}, \tag{4}$$

and

$$\forall \varepsilon > 0 \quad \exists K_\varepsilon > 0; \quad \forall u \in (0,1]; \quad \theta^0(u) \leq K_\varepsilon u^{-\frac{1}{2}-\varepsilon}. \tag{5}$$

Note that under Conditions (4) and (5), the fractional stochastic differential equation (3) admits a unique weak solution, see Theorem 2 and associated references.

We now give our result for the conditional laws of generalized rough Heston models (which Model (1) is a particular case of). Let $(S_t, V_t)_{t\geq 0}$ be defined by (3). We have the following theorem, proved in Section 5.1.

Theorem 1. *The law of the process* $(S_t^{t_0}, V_t^{t_0})_{t\geq 0}$ *is that of a generalized rough Heston model with the following dynamic:*

$$dS_t^{t_0} = S_t^{t_0}\sqrt{V_t^{t_0}}dW_t^{t_0}, \quad S_0^{t_0} = S_{t_0}$$

$$V_t^{t_0} = V_{t_0} + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\lambda(\theta^{t_0}(u) - V_u^{t_0})du + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\nu\sqrt{V_u^{t_0}}dB_u^{t_0},$$

with $(W_t^{t_0}, B_t^{t_0})_{t\geq 0} = (W_{t_0+t} - W_{t_0}, B_{t_0+t} - B_{t_0})_{t\geq 0}$ *a two-dimensional Brownian motion with correlation* ρ*, independent of* $\mathscr{F}_{t_0}$ *and*

$$\theta^{t_0}(u) = \theta^0(t_0+u) + \frac{\alpha}{\lambda\Gamma(1-\alpha)}\int_0^{t_0}(t_0 - v + u)^{-1-\alpha}(V_v - V_{t_0})dv + \frac{(u+t_0)^{-\alpha}}{\lambda\Gamma(1-\alpha)}(V_0 - V_{t_0}),$$

which is an $\mathscr{F}_{t_0}$*-measurable function continuous on* $\mathbb{R}_+^*$ *such that Conditions* (4) *and* (5) *(where the index* 0 *should be replaced by* t_0*) are satisfied.*

Hence the class of generalized rough Heston models is stable with respect to conditioning. The conditional law of a rough Heston model is still that of a rough Heston model. The only difference is a modification in the mean-reversion level function. In particular, when considering the usual rough Heston model (1), the constant parameter θ becomes an $\mathscr{F}_{t_0}$ measurable function when taking conditional law at time t_0. This result will be crucial to derive hedging strategies in the rough Heston framework, and more generally to understand the state variables associated to rough Heston type dynamics.

3 Characteristic function of generalized rough Heston models

The goal of this section is to derive the extended characteristic functions for the log-price in the rough Heston model (3). This together with Theorem 1 will enable us to derive conditional characteristic functions, leading to hedging strategies. The first step to achieve this goal is to build a suitable sequence of processes converging to the generalized rough Heston model of Definition 1. Then we will be able to do computations on these processes (notably deriving characteristic functions), and pass them to the limit to obtain results for generalized rough Heston models.

3.1 Generalized rough Heston models as limit of nearly unstable Hawkes processes

In Chapter II, a microscopic price model, based on two-dimensional Hawkes processes, is built so that it converges on the long run after suitable rescaling to a rough Heston log-price (with constant mean-reversion). Then, characteristic functions are obtained from this result. Such method could easily be extended to obtain a generalized rough Heston model in the limit. However, it would only enable us to compute (2) with $a = 0$. This is not enough so that classical Fourier inversion methods such as that in [CM99] can be rigorously applied to compute prices and hedging portfolios.

Thus we use another approach in this section, quite similar to that of [JR16b]. We consider a sequence of one-dimensional Hawkes processes $(N_t^T)_{t\geq 0}$, indexed by $T > 0$ going to infinity, with intensity given by

$$\lambda_t^T = \mu_T + \int_0^t a_T \varphi(t-s) dN_s^T,$$

where μ_T and a_T are positive constants with $a_T < 1$ and $\varphi : \mathbb{R}_+^* \to \mathbb{R}_+$ is integrable such that $\int_0^\infty \varphi = 1$. In [JR16b], it is shown that provided

$$x^\alpha \int_x^\infty \varphi(s) ds \underset{x\to\infty}{\longrightarrow} \frac{1}{\Gamma(1-\alpha)}, \quad \alpha \in (1/2, 1), \tag{6}$$

and

$$T^\alpha (1 - a_T) \underset{T\to\infty}{\longrightarrow} \lambda, \quad T^{1-\alpha}\mu_T \underset{T\to\infty}{\longrightarrow} \lambda/\nu^2,$$

for some positive constants λ and ν, a suitably rescaled version of the intensity process λ_t^T asymptotically behaves as the variance process of a rough Heston model with constant mean-reversion parameter such as (1) and with initial variance equal to zero. To obtain a time-dependent mean-reversion level and a non-zero starting value in the limit, we are inspired by an idea in Chapter II, where it is shown that a time-dependent μ_T is a way to modify some parameters in the limit. More precisely, we consider the following assumption, where $f^{\alpha,1}$ denotes the Mittag-Leffler density function defined in Appendix III.A.1.

Assumption 1. *There exist $\lambda, \nu > 0$, $\alpha \in (1/2, 1)$ and $V_0 > 0$ such that for $T > 1/\lambda^{-1/\alpha}$ and $t \geq 0$,*

$$\lambda_t^T = \mu_T \zeta^T(t) + \int_0^t \varphi^T(t-s) dN_s^T,$$

where

$$a_T = 1 - \lambda T^{-\alpha}, \quad \mu_T = (\lambda/\nu^2) T^{\alpha-1}, \quad \varphi^T = a_T \varphi,$$

with $\varphi = f^{\alpha,1}$ and

$$\zeta^T(t) = V_0 \Big(\frac{1}{1-a_T} (1 - \int_0^t \varphi^T(t-s) ds) - \int_0^t \varphi^T(t-u) du \Big) + \int_0^t \varphi^T(t-u) \theta^0(u/T) du,$$

where $\theta^0()$ satisfies the assumptions of Definition 1.

Note that we are working in the so-called nearly unstable case for Hawkes processes since the L^1 norm of the kernel φ^T converges to one. Furthermore remark that (6) is satisfied, see Appendix III.A.1.

Remark 1. *Remark that ζ^T can also be written as follows*

$$\zeta^T(t) = \int_0^t \varphi^T(t-u)\theta^0(u/T) du + V_0 \Big(\frac{T^\alpha}{\lambda} \int_t^\infty \varphi(s) ds + \lambda T^{-\alpha} \int_0^t \varphi(s) ds \Big).$$

Therefore using that $I^{1-\alpha}\varphi^T(t) = \int_t^\infty \varphi^T$, see Appendix III.A.1, together with Condition (4) *we get*

$$\begin{aligned}
\zeta^T(t) &\geq -\frac{V_0}{\lambda\Gamma(1-\alpha)} T^\alpha \int_0^t \varphi^T(t-u) u^{-\alpha} du + V_0 \Big(\frac{T^\alpha}{\lambda} \int_t^\infty \varphi(s) ds + \lambda T^{-\alpha} \int_0^t \varphi(s) ds \Big) \\
&= -\frac{V_0}{\lambda} T^\alpha \int_t^\infty \varphi^T(s) ds + V_0 \Big(\frac{T^\alpha}{\lambda} \int_t^\infty \varphi(s) ds + \lambda T^{-\alpha} \int_0^t \varphi(s) ds \Big) \\
&= V_0 \mu_T \Big(\int_t^\infty \varphi(s) ds + \lambda T^{-\alpha} \int_0^t \varphi(s) ds \Big).
\end{aligned}$$

This shows that ζ^T is a positive function and thus that the intensity process λ_t^T is well-defined.

We define $M_t^T = N_t^T - \int_0^t \lambda_s^T ds$ and

$$X_t^T = \nu^2 \frac{1-a_T}{T^\alpha \lambda} N_{tT}^T,\ \Lambda_t^T = \nu^2 \frac{1-a_T}{T^\alpha \lambda} \int_0^{tT} \lambda_s^T ds,\ Z_t^T = \nu \sqrt{\frac{1-a_T}{T^\alpha \lambda}} M_{tT}^T.$$

Conditions (4) and (5) on the function θ^0 allow us to adapt the proofs in [JR16b] and Chapter II in a straightforward way to obtain the following result.

Theorem 2. *Let $t_0 > 0$. As $T \to \infty$, under Assumption 1, the process $\big(\Lambda_t^T, X_t^T, Z_t^T\big)_{t\in[0,t_0]}$ converges in law for the Skorokhod topology to (Λ, X, Z), where*

- $\Lambda_t = X_t = \displaystyle\int_0^t V_s ds.$
- $Z_t = \displaystyle\int_0^t \sqrt{V_s} dB_s$ *which is a continuous martingale.*
- V *is the unique weak solution of the rough stochastic differential equation*

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda(\theta^0(s) - V_s) ds + \frac{\nu}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} dB_s,$$

where B is a Brownian motion. Furthermore, the process V is non-negative and has Hölder regularity $\alpha - 1/2 - \varepsilon$ for any $\varepsilon > 0$.

Theorem 2 will be one of the key results to obtain the extended characteristic function of the log-price in generalized rough Heston models.

3.2 Conditions for finite moments in generalized rough Heston models

Recall that we aim at computing (2) with $\Re(z) \neq 0$. A preliminary step towards this is to derive sufficient conditions for the finiteness of the moments of S_t and $\exp\big(\int_0^t V_s ds\big)$ in generalized rough Heston models. To obtain such result, we use Theorem 2. Let $a \in \mathbb{R}$. First, note that

$$(S_t)^a = (S_0)^a \exp\big(a\rho \int_0^t \sqrt{V_s} dB_s - \frac{a}{2} \int_0^t V_s ds + a\sqrt{1-\rho^2} \int_0^t \sqrt{V_s} dB_s^\perp\big).$$

Consequently, we have

$$\mathbb{E}[(S_t)^a] = (S_0)^a \mathbb{E}\big[\exp(a\rho \int_0^t \sqrt{V_s} dB_s + \frac{1}{2}(-a + a^2(1-\rho^2)) \int_0^t V_s ds)\big].$$

Now define

$$M_t = \exp\big(a\rho \int_0^t \sqrt{V_s} dB_s - \frac{a^2\rho^2}{2} \int_0^t V_s ds\big).$$

The process M_t is a positive local martingale and actually, by Proposition 3 in Appendix, a true martingale. Define the corresponding probability measure $\mathbb{Q}$:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}_t} = M_t.$$

By Girsanov theorem, under $\mathbb{Q}$,

$$B_t^{\mathbb{Q}} = B_t - a\rho\int_0^t \sqrt{V_s}ds$$

is a $\mathbb{F}$-Brownian motion. Consequently, under $\mathbb{Q}$, V defined in (3) is still the variance process of a generalized rough Heston model, but with different parameters:

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\tilde{\lambda}(\tilde{\theta^0}(u) - V_u)du + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\nu\sqrt{V_u}dB_u^{\mathbb{Q}}, \tag{7}$$

where

$$\tilde{\lambda} = \lambda - \rho\nu a, \quad \tilde{\theta^0}(t) = \frac{\lambda\theta^0(t)}{\lambda-\rho\nu a},$$

provided that $\lambda - \rho\nu a > 0$. Hence we obtain

$$\mathbb{E}[(S_t)^a] = (S_0)^a\mathbb{E}_{\mathbb{Q}}[\exp(\frac{1}{2}(-a+a^2)\int_0^t V_s ds)]. \tag{8}$$

Therefore, a sufficient condition on a for

$$\mathbb{E}_{\mathbb{Q}}[\exp(\frac{1}{2}(-a+a^2)\int_0^t V_s ds)] < \infty \tag{9}$$

will readily imply a sufficient condition for the finiteness of $\mathbb{E}[(S_t)^a]$.

We now explain how to derive such condition. Recall that from Theorem 2, $\nu^2 T^{-2\alpha} N_{tT}^T$ converges in law to $\int_0^t V_s ds$. Thus we look first for a condition on $a \in \mathbb{R}$ for which

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N_{tT}^T)] < \infty, \tag{10}$$

for large enough $T > 0$ and fixed $t > 0$. This is done using a population interpretation of Hawkes processes, see Appendix III.C.1. It leads us to a sufficient condition on $a \in \mathbb{R}$ for (9). Furthermore, we are able to compute explicitly the expectation in (10), see Appendix III.C.1. Thus we can pass to the limit as T goes to infinity and then obtain an explicit expression for the expectation in (9). More precisely, we have the following result whose proof is given in Section 5.2, where $a_0(t)$ is defined for $t > 0$ by

$$a_0(t) = \frac{1}{2\nu^2}(\lambda + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)})^2.$$

Theorem 3. *Let V be the variance process of the generalized rough Heston model* (3)*. For any $t > 0$ and $a < a_0(t)$,*

$$\mathbb{E}\big[\exp(a\int_0^t V_s ds)\big] < \infty$$

and

$$\mathbb{E}\big[\exp(a\int_0^t V_s ds)\big] = \exp\Big(\int_0^t g(a, t-s)(\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\Big),$$

where $g(a,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha g(a,s) = a - \lambda g(a,s) + \frac{\nu^2}{2} g(a,s)^2, \quad s \le t, \quad I^{1-\alpha} g(a,0) = 0.$$

For any $0 \le s \le t$, this function satisfies

$$g(a,s) \le \frac{c}{\nu^2}\Big(\frac{\alpha s^{-\alpha}}{\Gamma(1-\alpha)} + \nu\sqrt{a_0(s) - a}\Big)$$

for some constant $c > 0$. Furthermore, for fixed $0 \le s \le t$, $a \to g(a,s)$ is non-decreasing and $s \to g(a,s)$ is non-increasing on $[0,t]$ if $a < 0$ and non-decreasing if $a > 0$.

Let S_t denote the price in the generalized rough Heston model of Definition 1. Using (8), we obtain the following corollary on the moments of S_t.

Corollary 1. *Let $t > 0$. Assume*

$$\lambda - \rho\nu a > 0, \quad a_-(t) < a < a_+(t),$$

where

$$a_-(t) = \frac{\nu^2 - 2\rho\nu X(t) + \sqrt{\Delta(t)}}{2\nu^2(1-\rho^2)}, \quad a_+(t) = \frac{\nu^2 - 2\rho\nu X(t) - \sqrt{\Delta(t)}}{2\nu^2(1-\rho^2)},$$

with

$$X(t) = \lambda + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)}, \quad \Delta(t) = 4\nu^2 X(t)^2 + \nu^4 - 4\rho\nu^3 X(t).$$

Then we have

$$\mathbb{E}[(S_t)^a] < \infty.$$

Furthermore,

$$\mathbb{E}[(S_t)^a] = (S_0)^a \exp\Big(\int_0^t h(a, t-s)(\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\Big),$$

where $h(a,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha h(a,s) = \frac{a^2 - a}{2} - (\lambda - \rho\nu a)h(a,s) + \frac{\nu^2}{2} h(a,s)^2, \quad s \le t, \quad I^{1-\alpha} h(a,0) = 0.$$

Remark 2. *Note that if we formally take $\alpha = 1$ in Corollary 1, our model coincides with the classical Heston model. In that case $X(t) = \lambda$ and therefore a_- and a_+ do not depend on t. Moreover the set of $a \in \mathbb{R}$ such that*

$$\lambda - \rho \nu a > 0, \quad a_- \leq a \leq a_+,$$

exactly corresponds to that of $a \in \mathbb{R}$ for which

$$\forall t \geq 0, \quad \mathbb{E}[(S_t)^a] < \infty,$$

see [AP07] for further details on moment explosions for the classical Heston model.

Proof of Corollary 1:

Recall that from (8),

$$\mathbb{E}[(S_t)^a] = (S_0)^a \mathbb{E}_{\mathbb{Q}}[\exp(\frac{1}{2}(-a + a^2)\int_0^t V_s ds)].$$

From Theorem 3 and the fact that under $\mathbb{Q}$, V follows (7), this quantity is finite if $\lambda - \rho\nu a > 0$ and

$$\frac{1}{2}(-a + a^2) < \tilde{a}_0(t) = \frac{1}{2\nu^2}(\tilde{\lambda} + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)})^2 = \frac{1}{2\nu^2}(\lambda - \rho\nu a + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)})^2.$$

This is equivalent to

$$a^2\nu^2(1-\rho^2) + a(-\nu^2 + 2X(t)\rho\nu) - X(t)^2 < 0.$$

The conditions on $a \in \mathbb{R}$ stated in Corollary 1 follow. Finally, the expression of $\mathbb{E}[(S_t)^a]$ is easily obtained using (8) together with Theorem 3. □

3.3 Characteristic functions of generalized rough Heston models

We are now ready to derive the characteristic functions of generalized rough Heston models. Let $t > 0$. We want to compute

$$R(z,t) = \mathbb{E}\big[\exp\big(z\log(S_t/S_0)\big)\big],$$

where $z \in \mathbb{C}$ satisfies

$$z = a + ib, \quad a, b \in \mathbb{R}, \quad \lambda - \rho\nu a > 0, \quad a_-(t) < a < a_+(t), \tag{11}$$

where $a_-(t)$ and $a_+(t)$ are defined in Corollary 1. Recall that from Corollary 1, (11) implies that $\exp\big(z\log(S_t/S_0)\big)$ is integrable and therefore $R(z,t)$ is well-defined.

Using the same computations as in the preceding sections, we get

$$R(z,t) = \mathbb{E}_{\mathbb{Q}}\big[\exp\big(ib\rho\int_0^t \sqrt{V_s}dB_s^{\mathbb{Q}} + \frac{1}{2}(\rho^2b^2 + z^2 - z)\int_0^t V_s ds\big)\big]. \tag{12}$$

As already seen, under $\mathbb{Q}$, V still follows the variance process of a generalized rough Heston model driven by the Brownian motion $B^{\mathbb{Q}}$, see (7). Thus, we need to study

$$G(z,x,t) = \mathbb{E}\big[\exp\big(ix\int_0^t \sqrt{V_s}dB_s + z\int_0^t V_s ds\big)\big],$$

with $x \in \mathbb{R}$, $z \in \mathbb{C}$ such that $\Re(z) < a_0(t)$, ($a_0(t)$ is defined in Theorem 3), and V is the variance process of a generalized rough Heston model. To do so, we use again Theorem 2. Indeed $(\nu^2 T^{-2\alpha} N^T_{tT}, \nu T^{-\alpha} M^T_{tT})$ converges in law as T goes to infinity to $(\int_0^t V_s ds, \int_0^t \sqrt{V_s}dB_s)$. Computing

$$\mathbb{E}[\exp\big(ix\nu T^{-\alpha} M^T_{tT} + z\nu^2 T^{-2\alpha} N^T_{tT}\big)]$$

and passing to the limit, we obtain the following result whose proof is given in Section 5.3.

Theorem 4. *Let V be the variance process of the generalized rough Heston model* (3). *For any $t > 0$, $b \in \mathbb{R}$ and $z \in \mathbb{C}$ such that $\Re(z) < a_0(t)$,*

$$G(z,x,t) = \exp\big(\int_0^t \xi(z,x,t-s)(\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\big),$$

where $\xi(z,x,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha \xi(z,x,s) = z - \frac{x^2}{2} + (ix\nu - \lambda)\xi(z,x,s) + \frac{\nu^2}{2}\xi(z,x,s)^2, \quad s \le t, \quad I^{1-\alpha}\xi(z,x,0) = 0.$$

The following corollary is readily obtained from Theorem 4 together with (12).

Corollary 2. *Let $t > 0$ and $z \in \mathbb{C}$ satisfying* (11). *We have*

$$R(z,t) = \exp\big(\int_0^t h(z,t-s)(\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\big),$$

where $h(z,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha h(z,s) = \frac{1}{2}(z^2 - z) + (z\rho\nu - \lambda)h(z,s) + \frac{\nu^2}{2}h(z,s)^2, \quad s \le t, \quad I^{1-\alpha}h(z,0) = 0.$$

3.4 Connection with the forward variance curve

We now show how the characteristic function given in Corollary 2 can be written as a functional of the forward variance curve $(\mathbb{E}[V_t])_{t\ge 0}$. This property will be crucial in the next section when computing hedging portfolios. We first remark that the time-dependent parameter θ^0 can be directly linked to the forward variance curve through the following result.

Proposition 1. *Let V be the variance process of the generalized rough Heston model* (3). *For any $t \ge 0$, we have*

$$\mathbb{E}[V_t] = V_0\big(1 - F^{\alpha,\lambda}(t)\big) + \int_0^t f^{\alpha,\lambda}(t-s)\theta^0(s)ds, \tag{13}$$

where $F^{\alpha,\lambda}$ and $f^{\alpha,\lambda}$ are defined in Appendix III.A.1. Furthermore, θ^0 can be written as a functional of the forward variance curve as follows:

$$\lambda\theta^0(t) + V_0\frac{t^{-\alpha}}{\Gamma(1-\alpha)} = D^\alpha\mathbb{E}[V_t] + \lambda\mathbb{E}[V_t], \quad t > 0. \tag{14}$$

Proof of Proposition 1:

In the same way as in [JR16b], we can show that for any $t \geq 0$,

$$\mathbb{E}[\int_0^t V_s ds] < \infty.$$

So we have that $t \to \mathbb{E}[V_t]$ is locally integrable. Moreover $f^{\alpha,\lambda}$ is square-integrable, see Appendix III.A.1. Thus we obtain that for any $t \geq 0$,

$$\int_0^t f^{\alpha,\lambda}(t-s)^2 \mathbb{E}[V_s] ds < \infty.$$

Therefore,

$$\mathbb{E}[\int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s} dB_s] = 0.$$

Writing the dynamic of V under the following form as in [JR16b]:

$$V_t = V_0\big(1 - F^{\alpha,\lambda}(t)\big) + \int_0^t f^{\alpha,\lambda}(t-s)\theta^0(s)ds + \frac{\nu}{\lambda}\int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s}dB_s, \tag{15}$$

we deduce (13). Now using Fubini theorem and noting that $I^{1-\alpha} f^{\alpha,\lambda} = \lambda(1 - F^{\alpha,\lambda})$, see Appendix III.A.1, we get that for any $t \geq 0$,

$$I^{1-\alpha}\mathbb{E}[V_t] = V_0 \frac{t^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} + \int_0^t \lambda\big(1 - F^{\alpha,\lambda}(t-s)\big)(\theta^0(s) - V_0)ds.$$

Using Fubini again, this can be rewritten

$$I^{1-\alpha}\mathbb{E}[V_t] = V_0 \frac{t^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} + \int_0^t \lambda(\theta^0(s) - V_0)ds - \lambda \int_0^t \int_0^s f^{\alpha,\lambda}(s-u)(\theta^0(u) - V_0)duds.$$

Then from (13) we derive

$$I^{1-\alpha}\mathbb{E}[V_t] = V_0 \frac{t^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} + \int_0^t \lambda(\theta^0(s) - V_0)ds - \lambda \int_0^t (\mathbb{E}[V_s] - V_0)ds.$$

We finally obtain (14) by differentiating this last equality.

□

Remark 3. *Assume that the forward variance curve $t \to \mathbb{E}[V_t]$ is observed on the market through the implied volatility surface or liquid variance swaps, and that this curve admits a fractional derivative of order α. Then the mean-reversion function θ^0 can be chosen so that the model is consistent with this market forward variance curve by taking*

$$\lambda\theta^0(t) = D^\alpha(\mathbb{E}[V_.] - V_0)(t) + \lambda\mathbb{E}[V_t].$$

From Corollary 2 together with Proposition 1, we can eventually write the characteristic function of the log-price as a functional of the forward variance curve. Thus, it indicates that the forward variance curve is a relevant state variable in generalized rough Heston models. Such type of phenomena also appears in the class of models developed in [Ber05]. More precisely, we have the following corollary.

Corollary 3. *Let $t > 0$ and $z \in \mathbb{C}$ satisfying* (11). *We have*

$$R(z,t) = \exp\Big(\int_0^t \chi(z,t-s)\mathbb{E}[V_s]ds\Big),$$

where

$$\chi(z,t) = \frac{1}{2}(z^2 - z) + z\rho\nu h(z,t) + \frac{\nu^2}{2}h(z,t)^2,$$

with $h(z,.)$ the unique continuous solution of the fractional Riccati equation given in Corollary 2.

Thus, characteristic functions, and therefore conditional characteristic functions of the log-price can be written in term of the forward variance curve. This shows that this object plays the role of state variable in this infinite dimensional fractional setting. Actually, this result could probably be understood in a more general framework of affine processes, see [AJLP17, CT18].

Proof of Corollary 3:

By Lemma 2 in Appendix, for any $0 \le s \le t$,

$$h(z,s) = \int_0^s \frac{1}{\lambda} f^{\alpha,\lambda}(s-u)\chi(z,u)du. \tag{16}$$

Moreover, from (13) together with the fact that $I^{1-\alpha}f^{\alpha,\lambda} = \lambda(1-F^{\alpha,\lambda})$, see Appendix III.A.1, we have

$$\mathbb{E}[V_s] = \int_0^s \frac{1}{\lambda} f^{\alpha,\lambda}(s-u)(\lambda\theta^0(u) + V_0\frac{u^{-\alpha}}{\Gamma(1-\alpha)})du.$$

Then, using Fubini theorem, we obtain

$$\int_0^t \chi(z,t-s)\mathbb{E}[V_s]ds = \int_0^t \Big(\int_0^{t-s} \frac{1}{\lambda} f^{\alpha,\lambda}(t-s-u)\chi(z,u)du\Big)(\lambda\theta^0(s) + V_0\frac{s^{-\alpha}}{\Gamma(1-\alpha)})ds$$

and therefore

$$\int_0^t \chi(z,t-s)\mathbb{E}[V_s]ds = \int_0^t h(z,t-s)(\lambda\theta^0(s) + V_0\frac{s^{-\alpha}}{\Gamma(1-\alpha)})ds.$$

The result follows from Corollary 2. $\square$

4 Hedging under generalized rough Heston models

We consider a generalized rough Heston model with the additional assumption that $\rho \leq 0$. We show in this section how to compute explicitly hedging portfolios for vanilla options in such model. We treat here the case of a European call option with maturity $T > 0$ and strike $K > 0$. Nevertheless, the approach can be easily extended to other vanilla payoffs.

It is easy to see that we can find $a > 1$ such that the conditions of Corollary 1 are satisfied for any $t \geq 0$. Therefore, for any $t \geq 0$,

$$\mathbb{E}[(S_t)^a] < \infty.$$

We define the call option price process

$$C_t = \mathbb{E}[(S_T - K)_+ | \mathscr{F}_t], \quad 0 \leq t \leq T.$$

We write

$$X_t = \log(S_t), \quad t \geq 0$$

and

$$g(x) = e^{-ax}(e^x - K)_+, \quad x \in \mathbb{R}.$$

We have $g \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ and therefore

$$g(x) = \frac{1}{2\pi} \int_{b \in \mathbb{R}} \hat{g}(-b) e^{ibx} db,$$

where $\hat{g} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ is the Fourier transform of g. Note that we are able to compute explicitly $\hat{g}$:

$$\hat{g}(b) = \frac{e^{(1-a+ib)\log(K)}}{(ib-a)(ib-a+1)}, \quad b \in \mathbb{R}.$$

We then deduce by Fubini theorem that

$$C_t = \mathbb{E}[g(X_T) e^{aX_T} | \mathscr{F}_t] = \frac{1}{2\pi} \int_{b \in \mathbb{R}} \hat{g}(-b) P_t^T(a+ib) db, \tag{17}$$

where

$$P_t^T(a+ib) = \mathbb{E}[\exp\big((a+ib)X_T\big) | \mathscr{F}_t].$$

Using the fact that conditional on $\mathscr{F}_t$, S still follows a generalized rough Heston dynamic together with Corollary 3, we obtain

$$\mathbb{E}[\exp\big((a+ib)\log(S_T/S_t) | \mathscr{F}_t\big)] = \exp\Big(\int_0^{T-t} \chi(a+ib, T-t-s) \mathbb{E}[V_{s+t} | \mathscr{F}_t] ds\Big),$$

where χ is defined in Corollary 3. Thus,

$$P_t^T(a+ib) = \exp\Big((a+ib)\log(S_t) + \int_0^{T-t} \chi(a+ib, T-t-s) \mathbb{E}[V_{s+t} | \mathscr{F}_t] ds\Big). \tag{18}$$

Hence, from (18), we deduce that $P_t^T(a+ib)$ is a deterministic functional of the underlying spot price S_t and the forward variance curve until maturity T: $\mathbb{E}[V_{t+u}|\mathscr{F}_t], \quad 0 \le u \le T-t$.

Let

$$\mathcal{V}_{\alpha,\lambda} = \{\xi : \mathbb{R}_+ \to \mathbb{R}_+ \ , \quad \xi(t) = \int_0^t \frac{s^{-\alpha}}{\lambda\Gamma(1-\alpha)} f^{\alpha,\lambda}(t-s)\theta_\xi(s)ds, \quad \theta_\xi \text{ is continuous on } \mathbb{R}_+\}.$$

The space $\mathcal{V}_{\alpha,\lambda}$ is a metric space containing

$$\mathcal{V}^+_{\alpha,\lambda} = \{\xi \in \mathcal{V}_{\alpha,\lambda}, \quad \theta_\xi > 0 \text{ and for any } t>0, \quad \theta_\xi(t) = \xi(0) + t^\alpha \lambda\Gamma(1-\alpha)\theta^0_\xi(t), \quad \theta^0_\xi \text{ satisfies (5) }\},$$

which is the set of all possible forward variance curves produced by generalized rough Heston models. Note that from the same computations as for Proposition 1, we get the uniqueness of the function θ_ξ for each $\xi \in \mathcal{V}_{\alpha,\lambda}$ since we have

$$\theta_\xi(t) = \big(D^\alpha \xi(t) + \lambda\xi(t)\big)\Gamma(1-\alpha)t^\alpha, \quad t>0.$$

We equip $\mathcal{V}_{\alpha,\lambda}$ with the following complete metric:

$$d_{\alpha,\lambda}(\xi,\zeta) = \||\theta_\xi - \theta_\zeta| \wedge 1\|_\infty.$$

From (17) and (18), we get that the spot price and the forward variance curve are the relevant state variables for the call price process. Indeed, there exists a deterministic functional $C : \mathbb{R}_+ \times \mathbb{R}^*_+ \times \mathcal{V}_{\alpha,\lambda} \to \mathbb{R}$ such that

$$C_t = C\big(T-t, S_t, (\mathbb{E}[V_{s+t}|\mathscr{F}_t])_{s\ge 0}\big), \quad t \in [0,T],$$

where for any $t \ge 0$, $S \in \mathbb{R}_+$ and $\xi \in \mathcal{V}_{\alpha,\lambda}$

$$C\big(t,S,\xi\big) = \frac{1}{2\pi}\int_{b\in\mathbb{R}} \hat{g}(-b)L(a+ib,t,S,\xi)db, \tag{19}$$

with

$$L(a+ib,t,S,\xi) = \exp\big((a+ib)\log(S) + \int_0^t \chi(a+ib,t-s)\xi(s)ds\big).$$

In the following proposition, proved in Section 5.4, we give some useful regularity properties of the functional C.

Proposition 2. *Let $\xi \in \mathcal{V}^+_{\alpha,\lambda}$, $S>0$, $t>0$ and assume $|\rho|<1$. The function $C(t,.,\xi)$ defined in* (19) *is differentiable in S and its derivative is given by*

$$\partial_S C\big(t,S,\xi\big) = \frac{1}{2\pi}\int_{b\in\mathbb{R}} \frac{a+ib}{S}\hat{g}(-b)L(a+ib,t,S,\xi)db.$$

Moreover, the function $C(t,S,.)$ is differentiable in the sense of Fréchet in ξ, with derivative such that for any $\zeta \in \mathcal{V}_{\alpha,\lambda}$,

$$\partial_V C\big(t,S,\xi).\zeta = \int_0^t \big(\frac{1}{2\pi}\int_{b\in\mathbb{R}} \hat{g}(-b)L(a+ib,t,S,\xi)\chi(a+ib,t-s)db\big)\zeta(s)ds.$$

We end this section by stating our result showing how one can build a hedging portfolio by trading the underlying and the forward variance curve.

Theorem 5. *For any time $t \in [0,T]$, we have*

$$C_t = C_0 + \int_0^t \partial_S C(T-u, S_u, \mathbb{E}[V_{.+u}|\mathscr{F}_u]) dS_u + \int_0^t \partial_V C(T-u, S_u, \mathbb{E}[V_{.+u}|\mathscr{F}_u]).(d\mathbb{E}[V_{.+u}|\mathscr{F}_u]),$$

where

$$\partial_V C(T-u, S_u, \mathbb{E}[V_{.+u}|\mathscr{F}_u]).(d\mathbb{E}[V_{.+u}|\mathscr{F}_u])$$

denotes

$$\int_0^{T-u} \Big(\frac{1}{2\pi}\int_{b\in\mathbb{R}} \hat{g}(-b)L(a+ib, T-u, S_u, \mathbb{E}[V_{.+u}|\mathscr{F}_u])\chi(a+ib, T-u-s)db\Big) d\mathbb{E}[V_{s+u}|\mathscr{F}_u]ds,$$

with $d\mathbb{E}[V_x|\mathscr{F}_u]$ the Ito differential at time u of the martingale $M_u = \mathbb{E}[V_x|\mathscr{F}_u]$, $u \le x$.

Remark 4. *We actually also show that*

$$d\mathbb{E}[V_{s+u}|\mathscr{F}_u] = \frac{1}{\lambda} f^{\alpha,\lambda}(s)\nu\sqrt{V_u}dB_u.$$

The proof of Theorem 5 is given in Section 5.5. This result shows that in an idealistic setting where the underlying asset and the forward variance curve can be traded (in continuous time), perfect replication can be obtained in generalized rough Heston models. Of course, in practice, this strategy will be discretized and one will use liquid variance swaps or European options instead of the forward variance curve.

Remark 5. *It is interesting to remark that the price function $C(t,S,\xi)$ is solution of a Feynman-Kac type path-dependent partial differential equation. Let us define the following derivative according to time $t>0$:*

$$\partial_t C(t,S,\xi) = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\big(C(t-\varepsilon, S, \xi_{\varepsilon+.}) - C(t,S,\xi)\big).$$

We easily have that $L(a+ib,t,S,\xi)$ is solution of the following path-dependent PDE:

$$0 = \partial_t L + \frac{1}{2}(S\sqrt{\xi_0})^2 \partial_S^2 L + \frac{1}{2}(\nu\sqrt{\xi_0})\partial_V^2 L.(\frac{1}{\lambda}f^{\alpha,\lambda}, \frac{1}{\lambda}f^{\alpha,\lambda}) + \rho(S\sqrt{\xi_0})(\nu\sqrt{\xi_0})\partial_{S,V}^2 L.(\frac{1}{\lambda}f^{\alpha,\lambda}),$$

with the initial condition $L(a+ib,0,S,\xi) = S^{a+ib}$.

As in Proposition 2, we can show that C is twice differentiable in S and in V (in the sense of Fréchet for V), and that $\partial_t C$ is well-defined. So we can deduce that C satisfies the same path-dependent PDE:

$$0 = \partial_t C + \frac{1}{2}(S\sqrt{\xi_0})^2 \partial_S^2 C + \frac{1}{2}(\nu\sqrt{\xi_0})\partial_V^2 C.(\frac{1}{\lambda}f^{\alpha,\lambda}, \frac{1}{\lambda}f^{\alpha,\lambda}) + \rho(S\sqrt{\xi_0})(\nu\sqrt{\xi_0})\partial_{S,V}^2 C.(\frac{1}{\lambda}f^{\alpha,\lambda}),$$

with the initial condition $C(0,S,\xi) = (S-K)_+$.

Note that $\partial_V^2 C.(\frac{1}{\lambda}f^{\alpha,\lambda}, \frac{1}{\lambda}f^{\alpha,\lambda})$ (resp. $\partial_{S,V}^2 C.(\frac{1}{\lambda}f^{\alpha,\lambda})$) is the second Fréchet derivative of C (resp. the first Fréchet derivative of $\partial_S C$) applied on $(\frac{1}{\lambda}f^{\alpha,\lambda}, \frac{1}{\lambda}f^{\alpha,\lambda})$ (resp. $\frac{1}{\lambda}f^{\alpha,\lambda}$) which is well-defined even though $\frac{1}{\lambda}f^{\alpha,\lambda}$ does not belong to the metric space $\mathcal{V}_{\alpha,\lambda}$.

5 Proofs

The notion of fractional integrals and derivatives are heavily used in the proofs. Notations, definitions and useful results related to them are given in Appendix III.A.

5.1 Proof of Theorem 1

Finding the dynamic of $V_t^{t_0}$ conditional on $\mathscr{F}_{t_0}$ Using stochastic Fubini theorem, we can show that $I^{1-\alpha}V$ is a semi-martingale and for $t>0$,

$$(I^{1-\alpha}V)_t = V_0\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)}ds + \int_0^t \lambda(\theta^0(s)-V_s)ds + \int_0^t \nu\sqrt{V_s}dB_s.$$

Therefore,

$$\frac{1}{\Gamma(1-\alpha)}\int_0^{t+t_0}(t+t_0-u)^{-\alpha}V_u du$$

is equal to

$$\frac{1}{\Gamma(1-\alpha)}\int_0^{t_0}(t_0-u)^{-\alpha}V_u du + V_0\int_{t_0}^{t+t_0}\frac{1}{\Gamma(1-\alpha)}u^{-\alpha}du + \int_{t_0}^{t+t_0}\lambda(\theta^0(u)-V_u)du + \int_{t_0}^{t+t_0}\nu\sqrt{V_u}dB_u.$$

Using a change of variable, this can be written

$$\frac{1}{\Gamma(1-\alpha)}\int_0^{t_0}(t_0-u)^{-\alpha}V_u du + V_0\int_0^t \frac{1}{\Gamma(1-\alpha)}(t_0+u)^{-\alpha}du + \int_0^t \lambda(\theta^0(u+t_0)-V_u^{t_0})du + \int_0^t \nu\sqrt{V_u^{t_0}}dB_u^{t_0},$$

where $(B_t^{t_0})_{t\ge 0} = (B_{t+t_0}-B_{t_0})_{t\ge 0}$ is a Brownian motion independent of $\mathscr{F}_{t_0}$. Moreover, remarking that

$$I^{1-\alpha}V_t^{t_0} = \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-u)^{-\alpha}V_u^{t_0}du = \frac{1}{\Gamma(1-\alpha)}\int_{t_0}^{t+t_0}(t+t_0-u)^{-\alpha}V_u du$$

is equal to

$$\frac{1}{\Gamma(1-\alpha)}\int_0^{t+t_0}(t+t_0-u)^{-\alpha}V_u du - \frac{1}{\Gamma(1-\alpha)}\int_0^{t_0}(t+t_0-u)^{-\alpha}V_u du$$

and that

$$\frac{1}{\Gamma(1-\alpha)}\big((t_0-u)^{-\alpha}-(t+t_0-u)^{-\alpha}\big) = \frac{\alpha}{\Gamma(1-\alpha)}\int_0^t (t_0-u+v)^{-1-\alpha}dv,$$

we derive

$$\begin{aligned}I^{1-\alpha}V_t^{t_0} = {}& \frac{\alpha}{\Gamma(1-\alpha)}\int_0^{t_0}\int_0^t (t_0-u+v)^{-1-\alpha}dv V_u du + \int_0^t \frac{1}{\Gamma(1-\alpha)}(t_0+u)^{-\alpha}du V_0\\ &+ \int_0^t \lambda(\theta^0(u+t_0)-V_u^{t_0})du + \int_0^t \nu\sqrt{V_u^{t_0}}dB_u^{t_0}.\end{aligned}$$

This can be written as follows:

$$V_{t_0}\frac{t^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)}+\int_0^t\lambda(\theta^{t_0}(u)-V_u^{t_0})du+\int_0^t\nu\sqrt{V_u^{t_0}}dB_u^{t_0}, \tag{20}$$

with $(\theta^{t_0}(u))_{u\geq 0}$ a function which is $\mathscr{F}_{t_0}$ measurable and defined by

$$\theta^{t_0}(u)=\theta^0(t_0+u)+\frac{\alpha}{\lambda\Gamma(1-\alpha)}\int_0^{t_0}(t_0-v+u)^{-1-\alpha}(V_v-V_{t_0})dv+\frac{(u+t_0)^{-\alpha}}{\lambda\Gamma(1-\alpha)}(V_0-V_{t_0}).$$

Properties of θ^{t_0} It is clear that θ^{t_0} is continuous on $\mathbb{R}_+^*$. Moreover it is easy to see that for any $u>0$:

$$\theta^{t_0}(u)=\theta^0(t_0+u)+\frac{\alpha}{\lambda\Gamma(1-\alpha)}\int_0^{t_0}(t_0-v+u)^{-1-\alpha}V_vdv+\frac{1}{\lambda\Gamma(1-\alpha)}(V_0(u+t_0)^{-\alpha}-V_{t_0}u^{-\alpha}).$$

Since V is a non-negative process and θ^0 satisfies (4), we obtain that θ^{t_0} also satisfies (4). Finally, for fixed $\varepsilon>0$, V being $\alpha-1/2-\varepsilon$ Hölder continuous, there exists for almost each $\omega\in\Omega$ a positive constant $c_\varepsilon(\omega)$ such that for any $x,y\in[0,t_0]$:

$$|V_x-V_y|\leq c_\varepsilon(\omega)|x-y|^{\alpha-1/2-\varepsilon}.$$

Thus by integration by parts, we obtain for any $u\in(0,t_0]$

$$\begin{aligned}|\int_0^{t_0}(t_0-v+u)^{-1-\alpha}(V_v-V_{t_0})dv|&\leq c_\varepsilon(\omega)\int_0^{t_0}(t_0-v+u)^{-1-\alpha}(t_0-v)^{\alpha-1/2-\varepsilon}dv\\&=c_\varepsilon(\omega)u^{-1/2-\varepsilon}\int_0^{t_0/u}(x+1)^{-1-\alpha}x^{\alpha-1/2-\varepsilon}dx\\&\leq c_\varepsilon(\omega)u^{-1/2-\varepsilon}\int_0^{\infty}(x+1)^{-1-\alpha}x^{\alpha-1/2-\varepsilon}dx.\end{aligned}$$

Thus θ^{t_0} satisfies Condition (5) almost surely.

End of the proof We end the proof noting that from (20) and stochastic Fubini Theorem we have that

$$\int_0^t V_s^{t_0}ds=\frac{1}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}I^{1-\alpha}V_s^{t_0}ds$$

is equal to

$$V_{t_0}t+\frac{1}{\Gamma(\alpha)}\int_0^t\int_0^s(s-u)^{\alpha-1}\lambda(\theta^{t_0}(u)-V_u^{t_0})duds+\frac{1}{\Gamma(\alpha)}\int_0^t\int_0^s(s-u)^{\alpha-1}\nu\sqrt{V_u^{t_0}}dB_u^{t_0}ds.$$

Hence by differentiating the previous equality, we conclude that the dynamic of (S^{t_0},V^{t_0}) is given by

$$S_t^{t_0}=S_{t_0}\exp\Big(\int_0^t\sqrt{V_u^{t_0}}dW_u^{t_0}-\frac{1}{2}\int_0^tV_u^{t_0}du\Big),$$

$$V_t^{t_0}=V_{t_0}+\frac{1}{\Gamma(\alpha)}\int_0^t(t-u)^{\alpha-1}\lambda(\theta^{t_0}(u)-V_u^{t_0})du+\frac{1}{\Gamma(\alpha)}\int_0^t(t-u)^{\alpha-1}\nu\sqrt{V_u^{t_0}}dB_u^{t_0},$$

where $(W_t^{t_0})_{t\geq0}=(W_{t+t_0}-W_{t_0})_{t\geq0}$ is a Brownian motion independent of $\mathscr{F}_{t_0}$ and with correlation ρ with B^{t_0}.

5.2 Proof of Theorem 3

We work here with the sequence of Hawkes processes N^T defined in Assumption 1. Recall that for $t \geq 0$, from Theorem 2, $\nu^2 T^{-2\alpha} N^T_{tT}$, converges in law as T goes to infinity to

$$\int_0^t V_s ds,$$

where V is solution of the fractional stochastic differential equation (3). A key step for the proof of Theorem 3 is to show that for suitable $a \in \mathbb{R}$,

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N^T_{tT})] \underset{T\to\infty}{\longrightarrow} \mathbb{E}[\exp(a\int_0^t V_s ds)]. \tag{21}$$

Applying (31) in Appendix III.C.1 on the Hawkes process N^T, we write

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N^T_{tT})] = \exp\Big(\int_0^t \lambda\zeta^T(T(t-s))g^T(a,s)ds\Big),$$

with

$$g^T(a,t) = \nu^{-2} T^{\alpha}\big(\exp(a\nu^2 T^{-2\alpha})\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N^{f,T}_{tT})] - 1\big),$$

where $N^{f,T}$ is the Hawkes process of children cluster (with migrant rate φ^T and kernel φ^T), see Appendix III.C.1 for details. Moreover from Lemma 5, $\lambda\zeta^T(Ts)$ converges pointwise as T goes to infinity to

$$\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)}, \quad 0 < s \leq t.$$

Therefore, it is left to study the convergence of the function g^T.

Uniform boundedness of g^T From now on c denotes a positive constant that may vary from line to line.

From (30) in Appendix III.C, for each $t > 0$,

$$g^T(a,t) < \infty \tag{22}$$

provided

$$a\nu^2 T^{-2\alpha} \leq \int_0^{tT} \varphi^T - 1 - \log(\int_0^{tT} \varphi^T).$$

Moreover note that from Appendix III.A.1,

$$T^{2\alpha}\Big(\int_0^{tT} \varphi^T - 1 - \log(\int_0^{tT} \varphi^T)\Big) \underset{T\to\infty}{\longrightarrow} \frac{1}{2}(\lambda + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)})^2.$$

Thus Property (22) is satisfied for large enough $T > T_0(a,t,\lambda,\nu)$ and $a < a_0(t)$ with

$$a_0(t) = \frac{1}{2\nu^2}(\lambda + \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)})^2.$$

Furthermore, as $N^{f,T} \le N^{\infty,T}$ (which is the Galton-Watson process defined in Appendix III.C.2), using (34) in Appendix III.C.2, we obtain

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N_{tT}^{f,T})] \le \sum_{n\ge 0} \frac{\nu_T(t)^n e^{-\nu_T(t)(n+1)}}{n!}(n+1)^{n-1} e^{a\nu^2 T^{-2\alpha} n},$$

with $\nu_T(t) = \int_0^{tT} \varphi^T$. It is also easy to see from (34) (by taking $a = 0$ and $\nu = 1$ in (34)) that

$$1 = \sum_{n\ge 0} \frac{e^{-(n+1)}}{n!}(n+1)^{n-1}.$$

Consequently, we obtain

$$\begin{aligned} g^T(a,t) &\le \nu^{-2} T^\alpha \sum_{n\ge 0} \frac{e^{-(n+1)}}{n!}(n+1)^{n-1}\big(\nu_T(t)^n e^{(1-\nu_T(t))(n+1)} e^{a\nu^2 T^{-2\alpha}(n+1)} - 1\big) \\ &= \frac{1}{\nu^2 \nu_T(t)} T^\alpha \sum_{n\ge 0} \frac{e^{-(n+1)}}{n!}(n+1)^{n-1}\big(e^{x_T(t)(n+1)} - \nu_T(t)\big), \end{aligned}$$

where

$$x_T(t) = 1 - \nu_T(t) + \log(\nu_T(t)) + a\nu^2 T^{-2\alpha},$$

which is non-positive for $T > T_0(a,t,\lambda,\nu)$. Therefore

$$g^T(a,t) \le \frac{1}{\nu^2 \nu_T(t)} T^\alpha (1 - \nu_T(t)).$$

Assume now $a \le 0$, we use again $N^{f,T} \le N^{\infty,T}$ and (34) to get

$$\nu^{-2} T^\alpha \sum_{n\ge 0} \frac{e^{-(n+1)}}{n!}(n+1)^{n-1}\big(e^{x_T(t)(n+1)} - 1\big) \le g^T(a,t) \le 0.$$

By Stirling formula,

$$\frac{e^{-(n+1)}}{n!}(n+1)^{n-1} \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi(n+1)^3}}.$$

Thus,

$$-c\frac{1}{\nu^2 \nu_T(t)} T^\alpha \sum_{n\ge 0} \frac{1}{(n+1)^{3/2}}(1 - e^{x_T(t)(n+1)}) \le g^T(a,t) \le 0.$$

We deduce from Lemma 7 that

$$-\frac{c}{\nu^2 \nu_T(t)}\sqrt{-T^{2\alpha} x_T(t)} \le g^T(a,t) \le 0.$$

Therefore, for any $a < a_0(t)$:

$$|g^T(a,t)| \le c\frac{1}{\nu^2 \nu_T(t)}\big(T^\alpha(1 - \nu_T(t)) + \sqrt{-T^{2\alpha} x_T(t)}\big).$$

Finally, note that

$$T^{\alpha}(1-\nu_T(t)) \to \frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)}$$

and

$$T^{2\alpha}x_T(t) \to \nu^2(a_0(t)-a),$$

as T goes to infinity. Eventually,

$$\limsup_{T\to\infty}|g^T(a,t)| \le c\nu^{-2}\big(\frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)}+\nu\sqrt{(a_0(t)-a)}\big). \tag{23}$$

Uniform convergence of g^T We now fix $t_0>0$ and $a<a_0(t_0)$. The function $t\to g^T(a,t)$ being monotone and such that $g^T(a,0)=0$, we have that for any $0\le t\le t_0$

$$|g^T(a,t)|\le|g^T(a,t_0)|.$$

Moreover, from the previous section, there exists $T_0(t_0,a,\lambda,\nu)>0$ such that

$$\sup_{T\ge T_0}|g^T(a,t_0)|<\infty.$$

Hence, $g^T(a,.)$ is uniformly bounded in $0\le t\le t_0$ and $T\ge T_0$. We now assume that $T\ge T_0$.

Applying (31) in Appendix III.C.1 on the Hawkes process $N^{f,T}$, we obtain that for any $t\in[0,t_0]$,

$$\nu^2T^{-\alpha}g^T(a,t)+1=\exp\big(\nu^2T^{-2\alpha}a+\nu^2T^{1-\alpha}\int_0^t\varphi^T(Ts)g^T(a,t-s)ds\big).$$

By taking the logarithm of the previous expression, we write

$$\nu^2T^{-2\alpha}a+\nu^2T^{1-\alpha}\int_0^t\varphi^T(Ts)g^T(a,t-s)ds=\nu^2T^{-\alpha}g^T(a,t)-\frac{\nu^4}{2}T^{-2\alpha}g^T(a,t)^2-\varepsilon_1^T(t),$$

where $|T^{3\alpha}\varepsilon_1^T|$ is uniformly bounded in $t\in[0,t_0]$ and $T\ge T_0$. Hence

$$g^T(a,t)=T\int_0^t\varphi^T(Ts)g^T(a,t-s)ds+aT^{-\alpha}+\frac{\nu^2}{2}T^{-\alpha}g^T(a,t)^2+\frac{T^{\alpha}}{\nu^2}\varepsilon_1^T(t).$$

Thanks to Lemma 1,

$$g^T(a,t)=aT^{1-\alpha}\int_0^t\psi^T(Ts)ds+\frac{\nu^2}{2}T^{1-\alpha}\int_0^t\psi^T(Ts)g^T(a,t-s)^2ds+\varepsilon_2^T(t),$$

where

$$\varepsilon_2^T(t)=aT^{-\alpha}+\frac{\nu^2}{2}T^{-\alpha}g^T(a,t)^2+\frac{T^{\alpha}}{\nu^2}\varepsilon_1^T(t)+\frac{T^{\alpha+1}}{\nu^2}\int_0^t\psi^T(Ts)\varepsilon_1^T(t-s)ds$$

and $\psi^T = \sum_{k\geq 1} (\varphi^T)^{*k}$[2]. Note that $T^\alpha \varepsilon_2^T$ is uniformly bounded in $t \in [0, t_0]$ and $T \geq T_0$. Recall also that using Laplace transform computations as in [JR16b], we get

$$\lambda T^{1-\alpha}\psi^T(Tt) = a_T f^{\alpha,\lambda}(t). \tag{24}$$

Thus

$$g^T(a,t) = \int_0^t \frac{1}{\lambda} f^{\alpha,\lambda}(t-s)(a + \frac{\nu^2}{2} g^T(a,s)^2)ds + \varepsilon^T(t),$$

with $\varepsilon^T(t) = \varepsilon_2^T(t) - T^{-\alpha}\int_0^t f^{\alpha,\lambda}(t-s)(a+\frac{\nu^2}{2}g^T(a,s)^2)ds$. As done in the proof of Proposition 5 in Chapter II, using that $T^\alpha \varepsilon^T$ and $g^T(a,.)$ are uniformly bounded in $t \in [0, t_0]$ and $T \geq T_0$, together with Lemma 3, we deduce that $g^T(a,.)$ is a Cauchy sequence on $C([0,t_0],\mathbb{R})$. Therefore it converges to a continuous function $g(a,.)$ solution of the following equation:

$$g(a,t) = \int_0^t \frac{1}{\lambda} f^{\alpha,\lambda}(t-s)(a + \frac{\nu^2}{2} g(a,s)^2)ds.$$

By Lemma 2, it is equivalent to the fractional Riccati equation

$$D^\alpha g(a,t) = a - \lambda g(a,t) + \frac{\nu^2}{2} g(a,t)^2, \quad I^{1-\alpha} g(a,0) = 0,$$

which admits a unique continuous solution (the uniqueness being an obvious corollary of Lemma 3). Finally remark that from (23),

$$|g(a,t)| \leq \frac{c}{\nu^2}\big(\frac{\alpha t^{-\alpha}}{\Gamma(1-\alpha)} + \nu\sqrt{a_0(t) - a}\big).$$

Remark 6. *Note that for $a \geq 0$, $t \to g(-a,t)$ is non-increasing and since $g(-a,0) = 0$, we obtain the following inequality:*

$$g(-a,t) = \int_0^t \frac{1}{\lambda} f^{\alpha,\lambda}(t-s)(a + \frac{\nu^2}{2} g(-a,s)^2)ds \leq \frac{1}{\lambda} F^{\alpha,\lambda}(t)(-a + \frac{\nu^2}{2} g(-a,t)^2).$$

From this inequality, we get for $t > 0$

$$g(-a,t) \leq \frac{1 - \sqrt{1 + \frac{2\nu^2 a}{\lambda^2} F^{\alpha,\lambda}(t)^2}}{\frac{\nu^2}{\lambda} F^{\alpha,\lambda}(t)}.$$

End of the proof We know that for any $t \in [0, t_0]$ and for fixed $a < a_0(t_0)$,

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N_{tT}^T)] = \exp\big(\int_0^t \lambda\zeta^T(T(t-s)) g^T(a,s)ds\big).$$

Then, from the uniform convergence of $g^T(a,.)$ to $g(a,.)$ together with Lemma 4, Lemma 5 and the dominated convergence theorem, we obtain

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N_{tT}^T)] \to \exp\big(\int_0^t g(a,t-s)(\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)})ds\big)$$

[2]Recall that $(\varphi^T)^{*1} = \varphi$ and $(\varphi^T)^{*k}(t) = \int_0^t \varphi(t-s).(\varphi^T)^{*k-1}(s)ds$.

as T goes to infinity. By Fatou lemma, we deduce

$$\mathbb{E}[\exp\big(a\int_0^t V_s ds\big)] < \infty.$$

We end the proof by showing (21). The case $a \leq 0$ being obvious, we assume that $0 < a < a_0(t_0)$. Let $\varepsilon > 0$ such that $a(1+\varepsilon) < a_0(t_0)$. From the computations above, there exists $T_0(t, a, \lambda, \nu, \varepsilon)$ such that

$$\sup_{T \geq T_0} \mathbb{E}[\exp(a(1+\varepsilon)\nu^2 T^{-2\alpha} N_{tT}^T)] < \infty.$$

Therefore $\big(\exp(a\nu^2 T^{-2\alpha} N_{tT}^T)\big)_{T \geq T_0}$ is uniformly integrable and we conclude that

$$\mathbb{E}[\exp(a\nu^2 T^{-2\alpha} N_{tT}^T)] \to \mathbb{E}[\exp(a\int_0^t V_s ds)].$$

This ends the proof of Theorem 3.

5.3 Proof of Theorem 4

In this section, we place ourselves in the framework of generalized rough Heston models (3) and compute for $0 \leq t \leq t_0$

$$G(z, x, t) = \mathbb{E}[\exp\big(z\int_0^t V_s ds + ix\int_0^t \sqrt{V_s} dB_s\big)],$$

with $x \in \mathbb{R}$ and $z \in \mathbb{C}$ such that $\Re(z) < a_0(t)$, ($a_0(t)$ is defined in Theorem 3). It has been shown in the proof of Theorem 3 that there exists $T_0 > 0$ such that

$$\exp\big(z\nu^2 T^{-2\alpha} N_{tT}^T + ix\nu T^{-\alpha} M_{tT}^T\big)$$

is uniformly integrable for fixed t and $T \geq T_0$. We have that

$$\mathbb{E}[\exp\big(z\nu^2 T^{-2\alpha} N_{tT}^T + ix\nu T^{-\alpha} M_{tT}^T\big)] \tag{25}$$

is equal to

$$\mathbb{E}[\exp\big((z\nu^2 T^{-2\alpha} + ix\nu T^{-\alpha}) N_{tT}^T - ix\nu T^{-\alpha}\int_0^{tT}\int_0^s \varphi^T(s-u) dN_u^T ds - ix\nu T^{-\alpha}\mu_T \int_0^{tT} \zeta^T(s) ds\big)].$$

Let

$$f^T(t) = z\nu^2 T^{-2\alpha} + ix\nu T^{-\alpha} - ix\nu T^{-\alpha}\int_0^t \varphi^T(s) ds.$$

Using Fubini theorem, we get that (25) is also equal to

$$\mathbb{E}[\exp\big(\int_0^{tT} f^T(tT-s) dN_s^T - ix\frac{\lambda}{\nu}\int_0^t \zeta^T(sT) ds\big)].$$

Hence we deduce from Lemma 4 and Lemma 5 that

$$\begin{aligned} G(z, x, t) &= \lim_{T\to\infty} \mathbb{E}[\exp\big(z\nu^2 T^{-2\alpha} N_{tT}^T + ix\nu T^{-\alpha} M_{tT}^T\big)] \\ &= \exp\big(-\frac{ix}{\nu}\int_0^t \lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)} ds\big) \lim_{T\to\infty} \mathbb{E}[\exp\big(\int_0^{tT} f^T(tT-s) dN_s^T\big)]. \end{aligned}$$

Passing to the limit Applying (35) in Appendix III.C.3 on the Hawkes process N^T with the function f^T, we have that for large enough T,

$$\exp\Big(\int_0^{tT} f^T(tT-s)dN_s^T\Big)$$

is integrable and

$$\mathbb{E}[\exp\Big(\int_0^{tT} f^T(tT-s)dN_s^T\Big)] = \exp\Big(\int_0^t \lambda\zeta^T(T(t-s))k^T(z,x,s)ds\Big)$$

where

$$k^T(z,x,t) = \frac{1}{\nu^2}T^\alpha\big(e^{f^T(tT)}\mathbb{E}[e^{\int_0^{tT} f^T(tT-u)dN_u^{f,T}}]-1\big).$$

Furthermore, from Lemma 5, $\lambda\zeta^T(Ts)$ converges pointwise as T tends to infinity to

$$\lambda\theta^0(s) + \frac{V_0 s^{-\alpha}}{\Gamma(1-\alpha)}, \quad s \le t.$$

As in Section 5.2, we show the uniform boundedness of k^T and then its uniform convergence.

Uniform boundedness of k^T We start by noting that for $t \in [0,t_0]$,

$$\begin{aligned}|k^T(z,x,t)| \le \frac{1}{\nu^2}T^\alpha\big(&|e^{f^T(tT)}\mathbb{E}[e^{\int_0^{tT} f^T(tT-u)dN_u^{f,T}}] - e^{i\Im[f^T(tT)]}\mathbb{E}[e^{\int_0^{tT} i\Im[f^T(tT-u)]dN_u^{f,T}}]| \\ &+|e^{i\Im[f^T(tT)]}\mathbb{E}[e^{\int_0^{tT} i\Im[f^T(tT-u)]dN_u^{f,T}}]-1|\big).\end{aligned}$$

Using that $\Re[f^T] = \Re(z)\nu^2T^{-2\alpha}$ together with the following inequality

$$|e^{f^T(tT)}\mathbb{E}[e^{\int_0^{tT} f^T(tT-u)dN_u^{f,T}}]-e^{i\Im[f^T(tT)]}\mathbb{E}[e^{\int_0^{tT} i\Im[f^T(tT-u)]dN_u^{f,T}}]| \le |e^{\Re[f^T(tT)]}\mathbb{E}[e^{\int_0^{tT} \Re[f^T(tT-u)]dN_u^{f,T}}]-1|,$$

we derive

$$|k^T(z,x,t)| \le |k^T(\Re(z),0,t)| + |k^T(i\Im(z),x,t)|.$$

In Section 5.2, we have already shown that $k^T(\Re(z),0,t)$ is uniformly bounded in $t\in[0,t_0]$, for large enough T. It is now left to show the uniform boundedness of $k^T(i\Im(z),x,t)$. So now we take $z = ia$ where $a\in\mathbb{R}$. First, remark that f^T becomes

$$f^T(t) = ia\nu^2T^{-2\alpha} + ix\nu T^{-\alpha}(1-\int_0^t \varphi^T(s)ds) = i(a\nu^2+x\nu\lambda)T^{-2\alpha} + ix\nu T^{-\alpha}\int_t^\infty \varphi^T(s)ds.$$

We write

$$\tilde{X}_t^T = f^T(tT) + \int_0^{tT} f^T(tT-s)dN_s^{f,T}.$$

It is easy to see that

$$|f^T(tT)| \le |a|\nu^2T^{-2\alpha} + |x|\nu T^{-\alpha}.$$

Furthermore,

$$\mathbb{E}[\int_0^{tT} f^T(tT-s)dN_s^{f,T}]$$

is equal to

$$T^{-2\alpha}i(av^2+\lambda xv)\int_0^{tT}\mathbb{E}[\lambda_s^{f,T}]ds + ixvT^{-\alpha}\int_0^{tT}\big(\int_{tT-s}^{\infty}\varphi^T(u)du\big)\mathbb{E}[\lambda_s^{f,T}]ds,$$

where $\lambda^{f,T}$ is the intensity of the cluster of children Hawkes process $N^{f,T}$, see Appendix III.C.1. We recall its definition:

$$\lambda_u^{f,T} = \varphi^T(u) + \int_0^u \varphi^T(u-s)dN_s^{f,T}.$$

Using Lemma 1, we know that

$$\lambda_u^{f,T} = \psi^T(u) + \int_0^u \psi^T(u-s)dM_s^{f,T},$$

where $M^{f,T} = N^{f,T} - \int_0^{\cdot}\lambda_s^{f,T}ds$ is the martingale associated to $N^{f,T}$. Thanks to (24), we obtain

$$\mathbb{E}[\lambda_{tT}^{f,T}] = \frac{a_T f^{\alpha,\lambda}(t)}{\lambda}T^{\alpha-1}.$$

Therefore

$$\int_0^{tT}\mathbb{E}[\lambda_s^{f,T}]ds \le \frac{F^{\alpha,\lambda}(t)}{\lambda}T^{\alpha} \le \frac{F^{\alpha,\lambda}(t_0)}{\lambda}T^{\alpha} \le cT^{\alpha}.$$

Moreover, using that $y\in\mathbb{R}_+ \to y^{\alpha}\int_y^{\infty}\varphi^T$ is uniformly bounded in y and T and $I^{1-\alpha}f^{\alpha,\lambda} = \lambda(1-F^{\alpha,\lambda})$ (see Appendix III.A.1), we obtain

$$T\int_0^t\int_{T(t-s)}^{\infty}\varphi^T(u)du\mathbb{E}[\lambda_{sT}^{f,T}]ds \le \frac{c}{\lambda}\int_0^t (t-s)^{-\alpha}f^{\alpha,\lambda}(s)ds \le c.$$

We deduce then that

$$|\mathbb{E}[\tilde{X}_t^T]| \le cT^{-\alpha}(|a|v^2+|x|v).$$

Using that there exists $c>0$ such that for any $y\in\mathbb{R}$,

$$|e^{iy}-1-iy| \le cy^2,$$

we get

$$|\mathbb{E}[e^{\tilde{X}_t^T}-1]| \le c(|\mathbb{E}[\tilde{X}_t^T]| + \mathbb{E}[|\tilde{X}_t^T|^2]).$$

We have

$$\mathbb{E}[|\tilde{X}_t^T|^2] \le 2|f^T(t)|^2 + 2\mathbb{E}[|\int_0^{tT} f^T(tT-s)dN_s^{f,T}|^2]),$$

and

$$\mathbb{E}[|\int_0^{tT} f^T(tT-s)dN_s^{f,T}|^2] \le 2(\mathbb{E}[|\int_0^{tT} f^T(tT-s)dM_s^{f,T}|^2] + \mathbb{E}[|\int_0^{tT} f^T(tT-s)\lambda_s^{f,T}ds|^2]).$$

Since $\langle M^{f,T}, M^{f,T}\rangle = \int_0^{\cdot} \lambda^{f,T}(s)ds$, we obtain

$$\mathbb{E}[|\int_0^{tT} f^T(tT-s)dM_s^{f,T}|^2] = \mathbb{E}[\int_0^{tT} |f^T(tT-s)|^2 \lambda_s^{f,T} ds] \leq cT^{-\alpha}(|a|\nu^2+|x|\nu)^2.$$

Using Fubini theorem, we derive

$$\int_0^{tT} f^T(tT-s)\lambda_s^{f,T}ds = \int_0^{tT} f^T(tT-s)\psi^T(s)ds + \int_0^{tT}\int_0^{tT-s} f^T(tT-s-u)\psi^T(u)dudM_s^{f,T}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}[|\int_0^{tT} f^T(tT-s)\lambda_s^{f,T}ds|^2]\\
&\leq 2|\int_0^{tT} f^T(tT-s)\psi^T(s)ds|^2 + 2\int_0^{tT}|\int_0^{tT-s} f^T(tT-s-u)\psi^T(u)du|^2\mathbb{E}[\lambda_s^{f,T}]ds\\
&\leq c(|a|\nu^2+|x|\nu)^2T^{-2\alpha}(1+\int_0^{tT}\mathbb{E}[\lambda_s^{f,T}]ds)\\
&\leq c(|a|\nu^2+|x|\nu)^2T^{-\alpha}.
\end{aligned}$$

We eventually deduce

$$|k^T(ia,x,t)| \leq \frac{c}{\nu^2}\big(c(a,x)+c(a,x)^2\big), \quad c(a,x) = \nu^2|a|+\nu|x|.$$

End of the proof Using the same computations as in Section 5.2, we show that for fixed $z \in \mathbb{C}$ and $x \in \mathbb{R}$ such that $\Re(z) < a_0(t_0)$, $k^T(z,x,.)$ is a Cauchy sequence in $C([0,t_0],\mathbb{C})$ and therefore converges uniformly to $k(z,x,.)$ solution of

$$k(z,x,t) = ix\frac{1}{\nu} + \int_0^t \frac{1}{\lambda}f^{\alpha,\lambda}(t-s)\big(z+\frac{\nu^2}{2}k(z,x,t)^2\big)ds.$$

Therefore, we deduce

$$G(z,x,t) = \exp\Big(\int_0^t \xi(z,x,t-s)(\lambda\theta^0(s)+\frac{V_0s^{-\alpha}}{\Gamma(1-\alpha)})ds\Big)$$

where $\xi(z,x,t) = k(z,x,t) - ix/\nu$, which is solution of the following equation:

$$\xi(z,x,t) = \int_0^t \frac{1}{\lambda}f^{\alpha,\lambda}(t-s)\big(z-\frac{x^2}{2}+ib\nu\xi(z,x,s)+\frac{\nu^2}{2}\xi(z,x,s)^2\big)ds.$$

By Lemma 2, this is equivalent to the following fractional Riccati equation:

$$D^\alpha\xi(z,x,t) = z-\frac{x^2}{2}+(ix\nu-\lambda)\xi(z,x,t)+\frac{\nu^2}{2}\xi(z,x,t)^2, \quad I^{1-\alpha}\xi(z,x,0)=0.$$

We end this section with the following remarks which will be useful in the proof of Proposition 2.

Remark 7. *From the definition of k^T,*

$$\Re(k^T(z,x,t)) \leq k^T(\Re(z),0,t).$$

Passing to the limit as T goes to infinity, we get

$$\Re(\xi(z,x,t)) \leq \xi(\Re(z),0,t) = g(\Re(z),t),$$

with g defined in Theorem 3.

Remark 8. *From the proof of uniform boundedness of k^T and using the inequality for g in Theorem 3, we get that for any $t \in [0,t_0]$,*

$$|\xi(z,x,t)| \leq c(1+\sqrt{|\Re(z)|+\Im(z)^2+x^2}),$$

where c is a positive constant, $x \in \mathbb{R}$ and $z \in \mathbb{C}$ such that $\Re(z) < a_0(t_0)$.

5.4 Proof of Proposition 2

We fix $S>0$, $t>0$, $\xi \in \mathcal{V}^+_{\alpha,\lambda}$ and $a>1$ such that $\mathbb{E}[(S_t)^a] < \infty$. Using the same computations as in the proof of Corollary 3, we get

$$L(a+ib,t,S,\xi) = \exp\Big((a+ib)\log(S) + \int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} h(a+ib,t-s)\theta_\xi(s)ds\Big),$$

where h is the unique continuous solution of the fractional Riccati equation in Corollary 2. Moreover, thanks to Remark 7, we have that for any $t>0$ and $b \in \mathbb{R}$,

$$\Re\big(h(a+ib,s)\big) \leq q(b,s), \quad s \leq t,$$

where $q(b,.)$ is the unique continuous solution of the following fractional Riccati equation:

$$D^\alpha q(b,s) = \frac{a^2-a}{2} - (1-\rho^2)\frac{b^2}{2} - (\lambda-\rho\nu a)q(b,s) + \frac{\nu^2}{2}q(b,s)^2, \quad s \leq t, \quad I^{1-\alpha}q(b,0) = 0.$$

Note also that for large $|b|$, $\frac{a^2-a}{2} - (1-\rho^2)\frac{b^2}{2}$ is negative and therefore, using Remark 6,

$$q(b,s) \leq M(b,s) = \frac{1-\sqrt{1+\nu^2\frac{(1-\rho^2)b^2-(a^2-a)}{(\lambda-\rho\nu a)^2}F^{\alpha,\lambda-\rho\nu a}(s)^2}}{\frac{\nu^2}{\lambda-\rho\nu a}F^{\alpha,\lambda-\rho\nu a}(s)}, \quad s \leq t.$$

By dominated convergence theorem, we have

$$\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} M(b,t-s)\theta_\xi(s)ds \underset{b\to\infty}{\sim} -|b|\frac{\sqrt{1-\rho^2}}{\nu}\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)}\theta_\xi(s)ds.$$

Consequently there exists $c(t,\xi)>0$ such that for any $b \in \mathbb{R}$,

$$|L(a+ib,t,S,\xi)| \leq S^a \exp\big(-c(t,\xi)(-1+|b|)\big).$$

Moreover, it is easy to see that for any $b \in \mathbb{R}$, $L(a+ib,t,.,\xi)$ is differentiable in S and that

$$\partial_S L(a+ib,t,S,\xi) = \frac{a+ib}{S} L(a+ib,t,S,\xi).$$

Using (17) together with the dominated convergence theorem, we conclude that C is differentiable in the first variable S and that

$$\partial_S C(t,S,\xi) = \frac{1}{2\pi} \int_{b\in\mathbb{R}} \hat{g}(-b) \frac{a+ib}{S} L(a+ib,t,S,\xi) db.$$

Now let $\zeta \in \mathcal{V}_{\alpha,\lambda}$ and $\varepsilon_0 > 0$ such that $\theta_\xi(s) - \varepsilon_0 |\theta_\zeta(s)| > 0$ for any $s \in [0,t]$. We have that for any $\varepsilon \neq 0$, $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$,

$$\frac{1}{\varepsilon} |L(a+ib,t,S,\xi+\varepsilon\zeta) - L(a+ib,t,S,\xi)| \tag{26}$$

is equal to

$$S^a \exp\Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} \Re(h(a+ib,t-s))(\theta_\xi(s) - \varepsilon|\theta_\zeta(s)|) ds\Big)$$
$$\frac{1}{\varepsilon} |\exp\Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} \varepsilon h(a+ib,t-s)\theta_\zeta(s)_- ds\Big) - \exp\Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} \varepsilon h(a+ib,t-s)|\theta_\zeta(s)| ds\Big)|.$$

Recall that for large $|b|$, $\Re\big(h(a+ib,s)\big)$ is non-positive for any $s \le t$. Since there exists $c > 0$ such that for any $z, z' \in \mathbb{C}$ such that $\Re(z) \le 0$ and $\Re(z') \le 0$,

$$|\exp(z) - \exp(z')| \le c|z - z'|,$$

we conclude that (26) is dominated by

$$cS^a \Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} |h(a+ib,t-s)||\theta_\zeta(s)| ds\Big) \exp\Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} \Re(h(a+ib,t-s))(\theta_\xi(s) - \varepsilon_0|\theta_\zeta(s)|) ds\Big).$$

Using the same arguments as previously, we get that there exists $c(t,\xi,\zeta,\varepsilon_0) > 0$ such that

$$\exp\Big(\int_0^t \frac{s^{-\alpha}}{\Gamma(1-\alpha)} \Re(h(a+ib,t-s))(\theta_\xi(s) - \varepsilon_0|\theta_\zeta(s)|) ds\Big) \le \exp\big(-c(t,\xi,\zeta,\varepsilon_0)(-1+|b|)\big).$$

From Remark 8, we know that there exists $c(t) > 0$ such that for any $s \in [0,t]$ and $b \in \mathbb{R}$,

$$|h(a+ib,s)| \le c(t)(1+b^2).$$

Moreover, note that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} L(a+ib,t,S,\xi+\varepsilon\zeta) - L(a+ib,t,S,\xi)$$

is equal to

$$L(a+ib,t,S,\xi) \int_0^t \chi(a+ib,t-s)\zeta_s ds.$$

Consequently, by the dominated convergence theorem, $C(t,S,.)$ is differentiable in ξ in the direction of ζ in the Fréchet sense and

$$\partial_V C(t,S,\xi).\zeta = \frac{1}{2\pi} \int_{b\in\mathbb{R}} \hat{g}(-b) L(a+ib,t,S,\xi) \Big(\int_0^t \chi(a+ib,t-s)\zeta_s ds\Big) db.$$

5.5 Proof of Theorem 5

We first show that

$$\int_0^{T-t} \chi(a+ib,s)\mathbb{E}[V_{T-s}|\mathscr{F}_t]ds$$

is equal to

$$\int_0^T \chi(a+ib,s)\mathbb{E}[V_{T-s}]ds - \int_0^t \chi(a+ib,T-s)V_s ds + \int_0^t h(a+ib,T-s)\nu\sqrt{V_s}dB_s. \tag{27}$$

Recall that from Equation (15) we get

$$V_s = \mathbb{E}[V_s] + \int_0^s \frac{1}{\lambda} f^{\alpha,\lambda}(s-u)\nu\sqrt{V_u}dB_u.$$

This together with stochastic Fubini theorem give

$$\int_0^t \chi(a+ib,T-s)V_s ds = \int_0^t \chi(a+ib,T-s)\mathbb{E}[V_s]ds + \int_0^t \Big(\int_0^{t-u} \frac{1}{\lambda} f^{\alpha,\lambda}(s)\chi(a+ib,T-u-s)ds\Big)\nu\sqrt{V_u}dB_u.$$

We also have that for $s \in [0,T-t]$,

$$\mathbb{E}[V_{T-s}|\mathscr{F}_t] = \mathbb{E}[V_{T-s}] + \int_0^t \frac{1}{\lambda} f^{\alpha,\lambda}(T-s-u)\nu\sqrt{V_u}dB_u. \tag{28}$$

Then similarly,

$$\int_0^{T-t} \chi(a+ib,s)\mathbb{E}[V_{T-s}|\mathscr{F}_t]ds$$

is equal to

$$\int_0^{T-t} \chi(a+ib,s)\mathbb{E}[V_{T-s}]ds + \int_0^t \Big(\int_0^{T-t} \frac{1}{\lambda} f^{\alpha,\lambda}(T-s-u)\chi(a+ib,s)ds\Big)\nu\sqrt{V_u}dB_u.$$

This can also be written

$$\int_t^T \chi(a+ib,T-s)\mathbb{E}[V_s]ds + \int_0^t \Big(\int_{t-u}^{T-u} \frac{1}{\lambda} f^{\alpha,\lambda}(s)\chi(a+ib,T-u-s)ds\Big)\nu\sqrt{V_u}dB_u.$$

Finally we obtain that

$$\int_0^t \chi(a+ib,T-s)V_s ds + \int_0^{T-t} \chi(a+ib,s)\mathbb{E}[V_{T-s}|\mathscr{F}_t]ds$$

is equal to

$$\int_0^T \chi(a+ib,T-s)\mathbb{E}[V_s]ds + \int_0^t \Big(\int_0^{T-u} \frac{1}{\lambda} f^{\alpha,\lambda}(s)\chi(a+ib,T-u-s)ds\Big)\nu\sqrt{V_u}dB_u.$$

Thus (27) is directly deduced from the last relation and (16). Now using (27) together with Ito formula, we derive

$$P_t^T(a+ib) = P_0^T(a+ib) + \int_0^t (a+ib)P_s^T(a+ib)\sqrt{V_s}dW_s + \int_0^t P_s^T(a+ib)h(a+ib,T-s)\nu\sqrt{V_s}dB_s.$$

Then by (17) together with stochastic Fubini theorem and Proposition 2, we get

$$C_t = C_0 + \int_0^t \partial_S C(T-u, S_u, \mathbb{E}[V_{\cdot+u}|\mathscr{F}_u])dS_u + \frac{1}{2\pi}\int_0^t \Big(\int_{b\in\mathbb{R}} \hat{g}(-b)P_u^T(a+ib)h(a+ib,T-u)db\Big)\nu\sqrt{V_u}dB_u.$$

Furthermore, using again (16) together with Fubini theorem, we obtain that

$$\frac{1}{2\pi}\int_0^t \Big(\int_{b\in\mathbb{R}} \hat{g}(-b)P_u^T(a+ib)h(a+ib,T-u)db\Big)\nu\sqrt{V_u}dB_u$$

is equal to

$$\int_0^t \Big(\frac{1}{2\pi}\int_0^{T-u}\int_{b\in\mathbb{R}} \hat{g}(-b)P_u^T(a+ib)\chi(a+ib,T-u-s)dbds\Big)\frac{1}{\lambda}f^{\alpha,\lambda}(s)\nu\sqrt{V_u}dB_u.$$

This last quantity can be expressed in term of the forward variance curve thanks to (28).

III.A Fractional calculus

We define the fractional integral of order $r \in (0,1]$ of a function f as

$$I^r f(t) = \frac{1}{\Gamma(r)}\int_0^t (t-s)^{r-1}f(s)ds,$$

whenever the integral exists, and its the fractional derivative of order $r \in [0,1)$ as

$$D^r f(t) = \frac{1}{\Gamma(1-r)}\frac{d}{dt}\int_0^t (t-s)^{-r}f(s)ds,$$

whenever it exists.

We gather in this section some useful technical results related to fractional calculus.

III.A.1 Mittag-Leffler functions

Let $(\alpha,\beta) \in (\mathbb{R}_+^*)^2$. The Mittag-Leffler function $E_{\alpha,\beta}$ is defined and for $z \in \mathbb{C}$ by

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

For $(\alpha,\lambda) \in (0,1)\times\mathbb{R}_+$ we also define

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^\alpha), \ \ t>0,$$

$$F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s)ds, \ \ t \geq 0.$$

The function $f^{\alpha,\lambda}$ is a density function on $\mathbb{R}_+$ called Mittag-Leffler density function. The following properties of $f^{\alpha,\lambda}$ and $F^{\alpha,\lambda}$ can be found in [HMS11, Mai, MH08]. We have

$$f^{\alpha,\lambda}(t) \underset{t\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha)}t^{\alpha-1}, \ \ f^{\alpha,\lambda}(t) \underset{t\to\infty}{\sim} \frac{\alpha}{\lambda\Gamma(1-\alpha)}t^{-(\alpha+1)}$$

and

$$F^{\alpha,\lambda}(t) = 1 - E_{\alpha,1}(-\lambda t^{\alpha}),\ F^{\alpha,\lambda}(t) \underset{t\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha+1)} t^{\alpha},\ 1 - F^{\alpha,\lambda}(t) \underset{t\to\infty}{\sim} \frac{1}{\lambda\Gamma(1-\alpha)} t^{-\alpha}.$$

Note also that from obvious computations, we get $I^{1-\alpha} f^{\alpha,\lambda} = \lambda(1-F^{\alpha,\lambda})$. Finally, for $\alpha \in (1/2, 1)$, $f^{\alpha,\lambda}$ is square-integrable and its Laplace transform is given for $z \geq 0$ by

$$\hat{f}^{\alpha,\lambda}(z) = \int_0^{\infty} f_{\alpha,\lambda}(s) e^{-zs} ds = \frac{\lambda}{\lambda + z^{\alpha}}.$$

III.A.2 Wiener-Hopf equations

The following result enables us to solve Wiener-Hopf type equations, see for example [BDHM13b] for details.

Lemma 1. *Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^d$ and $\phi : \mathbb{R}_+ \to \mathcal{M}^d(\mathbb{R})$ be a matrix-valued function with integrable components such that the spectral radius of $\int_0^{\infty} \phi(s) ds$ is strictly smaller than 1. Then there exists a unique locally bounded function f from $\mathbb{R}$ to $\mathbb{R}^d$ solution of*

$$f(t) = g(t) + \int_0^t \phi(t-s).f(s) ds,\ t \geq 0$$

given by

$$f(t) = g(t) + \int_0^t \psi(t-s).g(s) ds,\ t \geq 0,$$

where $\psi = \sum_{k\geq 1} \phi^{*k}$[3].

III.A.3 Fractional differential equations

We now give some useful results about fractional differential equations. The next lemma can be found in [SKM93].

Lemma 2. *Let h be a continuous function from $[0,1]$ to $\mathbb{R}$, $\alpha \in (0,1]$ and $\lambda \in \mathbb{R}$. There is a unique solution to the equation*

$$D^{\alpha} y(t) = \lambda y(t) + h(t),\quad y(0) = 0$$

given by

$$y(t) = \int_0^t (t-s)^{\alpha-1} E_{\alpha,\alpha}\big(\lambda(t-s)^{\alpha}\big) h(s) ds.$$

We also have the following result whose proof can be found in Chapter II.

[3] Recall that $\phi^{*1} = \phi$ and $\phi^{*k}(t) = \int_0^t \phi(t-s).\phi^{*k-1}(s) ds$.

Lemma 3. *Let h be a non-negative continuous function from $[0,1]$ to $\mathbb{R}$ such that for any $t \in [0,1]$,*

$$h(t) \leq \varepsilon + C\int_0^t f^{\alpha,\lambda}(t-s)h(s)ds,$$

for some $\varepsilon \geq 0$ and $C \geq 0$. Then for any $t \in [0,1]$,

$$h(t) \leq C'\varepsilon,$$

with

$$C' = 1 + C\lambda\int_0^1 s^{\alpha-1}E_{\alpha,\alpha}\big(\lambda(C-1)s^{\alpha}\big)ds > 0.$$

In particular, if $\varepsilon = 0$ then $h = 0$.

III.A.4 Further results

Lemma 4. *There exists a positive constant c such that for any $T > 1/\lambda^{-1/\alpha}$ and $t \in (0,1)$:*

$$\zeta^T(tT) \leq c\Big(1 + \frac{t^{-\alpha}}{\Gamma(1-\alpha)}\Big).$$

Proof of Lemma 4:

Note that by Remark 1, we have

$$\zeta^T(tT) = \int_0^{tT}\varphi^T(tT-u)\theta^0(u/T)du + V_0\Big(\frac{T^{\alpha}}{\lambda}\int_{tT}^{\infty}\varphi(s)ds + \lambda T^{-\alpha}\int_0^{tT}\varphi(s)ds\Big).$$

Thanks to Appendix III.A.1, we have that for each $t \in (0,1]$:

$$T^{\alpha}\int_{tT}^{\infty}\varphi \leq ct^{-\alpha},$$

Moreover by using condition (5) and the fact that $\alpha > 1/2$, we write that for each $t \in (0,1]$:

$$\theta^0(t) \leq ct^{-\alpha}.$$

Thus:

$$\zeta^T(tT) \leq c\int_0^{tT}\varphi(Tt-u)(u/T)^{-\alpha}du + c(1+t^{-\alpha}).$$

Using Appendix III.A.1, we obtain

$$\int_0^{tT}\varphi(Tt-u)(u/T)^{-\alpha}du = \Gamma(1-\alpha)T^{\alpha}\int_{tT}^{\infty}\varphi \leq ct^{-\alpha},$$

which ends the proof. $\square$

Lemma 5. *For each $t \in (0,1]$, as T tends to infinity, $\zeta^T(tT)$ defined by Assumption 1 converges to*

$$V_0\frac{t^{-\alpha}}{\lambda\Gamma(1-\alpha)} + \theta^0(t).$$

Proof of Lemma 5:

Let $t > 0$. We have

$$\zeta^T(tT) = a_T \int_0^t T\varphi(T(t-s))\theta^0(s)ds + V_0\big(\frac{T^\alpha}{\lambda}\int_{tT}^\infty \varphi(s)ds + \lambda T^{-\alpha}\int_0^{tT}\varphi(s)ds\big).$$

Moreover, from Appendix III.A.1,

$$V_0\big(\frac{T^\alpha}{\lambda}\int_{tT}^\infty \varphi(s)ds + \lambda T^{-\alpha}\int_0^{tT}\varphi(s)ds\big)$$

converges to

$$V_0\frac{t^{-\alpha}}{\lambda\Gamma(1-\alpha)}.$$

Moreover, since θ^0 is continuous in t, for any $\varepsilon > 0$ there exists $\eta > 0$ such that for any $s \in [t-\eta, t]$,

$$|\theta^0(s) - \theta^0(t)| \le \varepsilon.$$

Hence from Appendix III.A.1 together with the fact that φ is non-increasing, we obtain

$$\begin{aligned}|\int_0^t T\varphi\big(T(t-s)\big)\big(\theta^0(s) - \theta^0(t)\big)ds| &\le \varepsilon\int_0^{T\eta}\varphi + \int_0^{t-\eta} T\varphi\big(T(t-s)\big)(|\theta^0(t)| + |\theta^0(s)|)ds\\ &\le \varepsilon + T\varphi(T\eta)\int_0^t (|\theta^0(t)| + |\theta^0(s)|)ds \le 2\varepsilon\end{aligned}$$

for large enough T. Thus $\int_0^t T\varphi(T(t-s))\theta^0(s)ds$ converges to $\theta^0(t)$.

□

Lemma 6. *If* $\theta^0 : (0,1] \to \mathbb{R}$ *satisfies Condition* (5)*, then for any* $0 < \varepsilon < \alpha - 1/2$*,*

$$t \to \int_0^t f^{\alpha,\lambda}(t-s)\theta^0(s)ds$$

has Hölder smoothness $\alpha - 1/2 - \varepsilon$ *on* $[0,1]$.

Proof of Lemma 6:

Using Proposition A.2 in [JR16b], we obtain that for any $\eta \in (0,\alpha)$,

$$\int_0^t f^{\alpha,\lambda}(t-s)\theta^0(s)ds = \int_0^t D^\eta f^{\alpha,\lambda}(t-s)I^\eta\theta^0(s).$$

Taking $\eta = 1/2 + \varepsilon$, we have that $I^\eta\theta^0$ is a bounded function. Then, using Proposition A.3 in [JR16b], we obtain that our function has Hölder regularity equal to $\alpha - \eta = \alpha - 1/2 - \varepsilon$.

□

Let $x \ge 0$. We define

$$S(x) = \sum_{n\ge 0}\frac{1}{(n+1)^{3/2}}(1 - e^{-x(n+1)}).$$

We have the following lemma.

Lemma 7. *There exists $c > 0$ such that for any $x \geq 0$:*

$$S(x) \leq c\sqrt{x}.$$

Proof of Lemma 7:
We have

$$S(x) = \sum_{n\geq 0} \frac{1}{(n+1)^{3/2}}(1-e^{-x}) \sum_{0\leq k\leq n} e^{-kx}.$$

This can be rewritten

$$S(x) = (1-e^{-x}) \sum_{k\geq 0} \xi_k e^{-kx},$$

with $\xi_k = \sum\limits_{n\geq k} \frac{1}{(n+1)^{3/2}}$, which is equivalent to $2/\sqrt{k+1}$ as k tends to infinity. Thus there exists $c > 0$ such that for any $x \geq 0$:

$$S(x) \leq c(1-e^{-x}) \sum_{k\geq 0} \frac{1}{\sqrt{k+1}} e^{-(k+1)x}.$$

We conclude using that

$$\sum_{k\geq 0} \frac{1}{\sqrt{k+1}} e^{-(k+1)x} \leq \sum_{k\geq 0} \int_k^{k+1} \frac{1}{\sqrt{y}} e^{-yx} dy = \frac{\Gamma(1/2)}{\sqrt{x}}$$

together with the fact that

$$1-e^{-x} \leq cx.$$

□

III.B Martingale property of the price in the generalized rough Heston model

Proposition 3. *The process S defined by the generalized rough Heston model in Definition 1 is a $\mathbb{F}$-martingale.*

Proof of Proposition 3:
Let $t_0 > 0$ such that $1/2 < a_0(t_0)$. Thanks to Theorem 3, Novikov's criterion holds:

$$\mathbb{E}[\exp(\frac{1}{2}\int_0^{t_0} V_s ds)] < \infty.$$

Therefore $(S_u)_{0\leq u\leq t_0}$ is a martingale and $\mathbb{E}[S_{t_0}] = S_0$.

Now, assume that for a given $n \in \mathbb{N}$, $\mathbb{E}[S_{nt_0}] = S_0$. Recall that conditional on $\mathscr{F}_{nt_0}$, the law of $(S_t^{nt_0}, V_t^{nt_0})_{t\geq 0} = (S_{t+nt_0}, V_{t+nt_0})_{t\geq 0}$ is still that of a rough Heston model with the following dynamic:

$$dS_t^{nt_0} = S_t^{nt_0}\sqrt{V_t^{nt_0}}dW_t^{nt_0}$$

$$V_t^{nt_0} = V_{nt_0} + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\lambda(\theta^{nt_0}(u) - V_u^{nt_0})du + \frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\nu\sqrt{V_u^{nt_0}}dB_u^{nt_0},$$

where θ^{nt_0} is a $\mathscr{F}_{nt_0}$-measurable function satisfying almost surely Conditions (4) and (5), and $(W^{nt_0}, B^{nt_0}) = (W_{.+nt_0} - W_{nt_0}, B_{.+nt_0} - B_{nt_0})$ is a Brownian motion independent of $\mathscr{F}_{nt_0}$. Since $1/2 < a_0(t_0)$, we have again Novikov's criterion

$$\mathbb{E}[\exp(\frac{1}{2}\int_0^{t_0} V_s^{nt_0} ds)|\mathscr{F}_{nt_0}] < \infty.$$

Therefore $\mathbb{E}[S_{t_0}^{nt_0}|\mathscr{F}_{nt_0}] = S_{nt_0}$ and so

$$\mathbb{E}[S_{(n+1)t_0}] = \mathbb{E}[S_{nt_0}] = S_0.$$

Consequently, for any $n \in \mathbb{N}$,

$$\mathbb{E}[S_{nt_0}] = S_0,$$

which ends the proof. □

III.C Moments properties for Hawkes processes

Here we consider a one-dimensional Hawkes process N with intensity

$$\lambda_t = \mu(t) + \int_0^t \varphi(t-s)dN_s,$$

such that $\mu,\ \varphi : \mathbb{R}_+ \to \mathbb{R}_+$ are locally integrable and $\int_0^\infty \varphi < 1$. We are interested in a sufficient condition on $a > 0$ so that

$$\mathbb{E}[e^{aN_t}] < \infty. \tag{29}$$

We will show that (29) holds provided

$$a \leq \int_0^t \varphi - 1 - \log(\int_0^t \varphi). \tag{30}$$

To do so, we recall the branching structure of Hawkes processes.

III.C.1 Branching structure of Hawkes processes

We recall that the Hawkes process N can be viewed as a population process in which migrants arrive according to a non-homogenous Poisson process N^0 with intensity μ. Each migrant gives birth to children according to a non-homogenous Poisson process with intensity φ and each child also gives birth to children according to non-homogenous Poisson process with the same intensity and so on.

Therefore, it is easy to see that the cluster of children created by a migrant has the law of a Hawkes process N^f with the same kernel function φ but with migrant rate φ. So, the intensity of N^f is given by :

$$\lambda_t^f = \varphi(t) + \int_0^t \varphi(t-s)dN_s^f.$$

Using the branching structure of the Hawkes process, we can see that we easily derive the following equality in law:

$$N_t = N_t^0 + \sum_{1\le k\le N_t^0} N_{t-T_k}^{f,k}$$

where the $(T_k)_{k\ge 1}$ are the arrival times of the migrants and $(N^{f,k})_{k\ge 1}$ are independent copies of N^f, independent of N^0. Then, we can show that for $a \ge 0$,

$$\mathbb{E}[e^{aN_t}] = \exp\Big(\int_0^t \mu(t-s)(e^a\mathbb{E}[e^{aN_s^f}]-1)ds\Big), \tag{31}$$

see Chapter II. This is smaller than

$$\exp\Big(\int_0^t \mu(s)ds(e^a\mathbb{E}[e^{aN_t^f}]-1)\Big).$$

Consequently, a sufficient condition to obtain (29) is

$$\mathbb{E}[e^{aN_t^f}] < \infty. \tag{32}$$

III.C.2 Galton-Watson structure and exponential moments

Let us consider now the Hawkes process N^f. Using the population interpretation given in the previous section on this process, N_t^f is the number of migrants and children arrived up to time t. Let $t > 0$. We define the process N^∞ from N^f as follows.

- We consider $N_t^{(0)}$ the number of migrants arrived up to time t, which is a Poisson variable with parameter $\nu = \int_0^t \varphi$.
- For each migrant arrived at time $T_k < t$, we consider the number of children of first generation made by the migrant during a period of time t, which is also a Poisson variable with parameter ν, independent of $N_t^{(0)}$. We denote by X_t^1 the set of all those children and $N_t^{(1)} = \#(X_t^1)$ their total number.
- For each child of n^{th} generation of the set X_t^n, we consider the number of its children that are made during a period of time t, which is also a Poisson variable with parameter ν, independent of the previous generations. We denote X_t^{n+1} the set of all those children and $N_t^{(n+1)} = \#(X_t^{n+1})$ their total number.

It is clear that $X_t = \bigcup_{n\ge 0} X_t^n$ contains all the individuals of the Hawkes process N^f arrived up to time t. So,

$$N_t^\infty = \#(X_t) = \sum_{n\ge 0} N_t^{(n)} \ge N_t^f.$$

Thus a sufficient condition to obtain (32) is

$$\mathbb{E}[e^{aN_t^\infty}] < \infty. \tag{33}$$

Now remark that $(N_t^{(n)})_{n\geq 0}$ is a Galton-Watson process. Indeed,

$$N_t^{(n+1)} = \sum_{1\leq k\leq N_t^{(n)}} \xi_{k,n+1}; \quad n \geq 0.$$

where $(\xi_{k,n})_{k,n\geq 1}$ are i.i.d Poisson random variables with parameter ν, independent of the $N_t^{(k)}$.

We classically have, see for example [Dwa69],

$$\mathbb{P}[N_t^{\infty} = n] = \frac{\nu^n e^{-\nu(n+1)}}{n!}(n+1)^{n-1}.$$

Consequently,

$$\mathbb{E}[e^{aN_t^{\infty}}] = \sum_{n\geq 0} \frac{\nu^n e^{-\nu(n+1)}}{n!}(n+1)^{n-1}e^{an}. \tag{34}$$

Using Stirling formula, we get

$$\frac{\nu^n e^{-\nu(n+1)}}{n!}(n+1)^{n-1}e^{an} \underset{n\to\infty}{\sim} \frac{(\nu e^{1-\nu+a})^n}{\sqrt{2\pi n^3}} e^{1-\nu}$$

Hence (33) holds if and only if

$$\nu e^{1-\nu+a} \leq 1,$$

which is equivalent to :

$$a \leq \int_0^t \varphi - 1 - \log(\int_0^t \varphi).$$

III.C.3 A useful equality

Let us consider $g : \mathbb{R}_+ \to \mathbb{R}$ continuous and $a \in \mathbb{R}$ satisfying (30). We know that $\exp(\int_0^t f(t-s)dN_s)$ is integrable, where $f = a + ig$. Using the branching structure of Hawkes processes presented in Appendix III.C.1, we deduce the following equality in law:

$$\int_0^t f(t-s)dN_s = \int_0^t f(t-s)dN_s^0 + \sum_{1\leq k\leq N_t^0} \int_0^{t-T_k} f(t-T_k-s)dN_s^{f,k}.$$

Therefore, we can show that

$$\mathbb{E}[\exp(\int_0^t f(t-s)dN_s)] = \exp\Big(\int_0^t \mu(t-s)(e^{f(s)}\mathbb{E}[e^{\int_0^s f(s-u)dN_u^f}] - 1)ds\Big). \tag{35}$$

CHAPTER IV

Markovian structure of the Volterra Heston model

Abstract

We characterize the Markovian and affine structure of the Volterra Heston model in terms of an infinite-dimensional adjusted forward process and specify its state space. More precisely, we show that it satisfies a stochastic partial differential equation and displays an exponentially-affine characteristic functional. As an application, we deduce an existence and uniqueness result for a Banach-space valued square-root process and provide its state space. This leads to another representation of the Volterra Heston model together with its Fourier-Laplace transform in terms of this possibly infinite system of affine diffusions.

Keywords: Affine Volterra processes, stochastic Volterra equations, Markovian representation, stochastic invariance, Riccati-Volterra equations, rough volatility.

1 Introduction

The Volterra Heston model is defined by the following dynamics

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0, \tag{1}$$

$$V_t = g_0(t) + \int_0^t K(t-s)\left(-\lambda V_s ds + \nu\sqrt{V_s}dW_s\right), \tag{2}$$

with $K \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$, $g_0 : \mathbb{R}_+ \to \mathbb{R}$, $\lambda, \nu \in \mathbb{R}_+$ and $B = \rho W + \sqrt{1-\rho^2}W^\perp$ such that $(W, W^\perp)$ is a two-dimensional Brownian motion and $\rho \in [-1, 1]$. It has been introduced in [AJLP17] for the purpose of financial modeling following the literature on so-called rough volatility models [GJR18]. Hence S_t typically represents a stock price at time t with instantaneous stochastic variance V_t.

This model nests as special cases the Heston model for $K \equiv 1$, and the rough Heston model of Chapter II, obtained by setting $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ for $\alpha \in (\frac{1}{2}, 1)$ and

$$g_0(t) = V_0 + \int_0^t K(s)\lambda\theta ds, \quad t \geq 0, \quad \text{for some } V_0, \theta \geq 0, \tag{3}$$

so that the only model parameters are $V_0, \theta, \lambda, \rho, \nu, \alpha$. Recall that the rough Heston model does not only fit remarkably well historical and implied volatilities of the market, but also enjoys a semi-closed formula for the characteristic function of the log-price in terms of a solution of a deterministic Riccati-Volterra integral equation.

In Chapter III, the authors highlight the crucial role of (3) in the design of hedging strategies for the rough Heston model. Here we consider more general input curves g_0. Our motivation is twofold. In practice, the function g_0 is intimately linked to the forward variance curve $(\mathbb{E}[V_t])_{t\geq 0}$. More precisely, taking the expectation in (2) leads to the following relation

$$\mathbb{E}[V_t] + \lambda \int_0^t K(t-s)\mathbb{E}[V_s] ds = g_0(t), \quad t \geq 0.$$

Thus, allowing for more general input curves g_0 leads to more consistency with the market forward variance curve. From a mathematical perspective, this enables us to understand the general picture behind the Markovian and affine nature of the Volterra Heston model (1)-(2).

More precisely, adapting the methods of [AJLP17], we provide a set of admissible input curves $\mathcal{G}_K$ defined in (9) such that (1)-(2) admits a unique $\mathbb{R}_+^2$-valued weak solution for any $g_0 \in \mathcal{G}_K$. In particular, we show that the Fourier-Laplace transform of $(\log S, V)$ is exponentially affine in $(\log S_0, g_0)$. Then we prove that, conditional on $\mathcal{F}_t$, the shifted Volterra Heston model $(S_{t+\cdot}, V_{t+\cdot})$ still has the same dynamics as in (1)-(2) provided that g_0 is replaced by the following adjusted forward process

$$g_t(x) = \mathbb{E}\left[V_{t+x} + \lambda \int_0^x K(x-s) V_{t+s} ds \;\middle|\; \mathcal{F}_t \right], \quad x \geq 0. \tag{4}$$

This leads to our main result which states that $\mathcal{G}_K$ is stochastically invariant with respect to the family $(g_t)_{t\geq 0}$. In other words, if we start from an initial admissible input curve $g_0 \in \mathcal{G}_K$, then g_t belongs to $\mathcal{G}_K$, for all $t \geq 0$, see Theorem 3. This in turn enables us to characterize the Markovian structure of (S, V) in terms of the stock price and the adjusted forward process $(g_t)_{t\geq 0}$. Furthermore, $(g_t)_{t\geq 0}$ can be realized as the unique $\mathcal{G}_K$-valued mild solution of the following stochastic partial differential equation of Heath–Jarrow–Morton-type

$$dg_t(x) = \left(\frac{d}{dx} g_t(x) - \lambda K(x) g_t(0) \right) dt + K(x)\nu\sqrt{g_t(0)} dW_t, \quad g_0 \in \mathcal{G}_K,$$

and displays an affine characteristic functional.

As an application, we establish the existence and uniqueness of a Banach-space valued square-root process and provide its state space. This leads to another representation of $(V_t, g_t)_{t\geq 0}$.

Moreover, the Fourier-Laplace transform of $(\log S, V)$ is shown to be an exponential affine functional of this process. These results are in the spirit of the Markovian representation of fractional Brownian motion, see [CC98, HS15].

The paper is organized as follows. In Section 2, we prove weak existence and uniqueness for the Volterra Heston model and provide its Fourier-Laplace transform. Section 3 characterizes the Markovian structure in terms of the adjusted forward variance process. Section 4 establishes the existence and uniqueness of a Banach-space valued square-root process and provides the link with the Volterra framework. In Appendix IV.A we derive general existence results for stochastic Volterra equations. Finally, for the convenience of the reader we recall in Appendix IV.B the framework and notations regarding stochastic convolutions as in [AJLP17].

Notations : Elements of $\mathbb{C}^m$ are viewed as column vectors, while elements of the dual space $(\mathbb{C}^m)^*$ are viewed as row vectors. For $h \geq 0$, Δ_h denotes the shift operator, i.e. $\Delta_h f(t) = f(t+h)$. If the function f on $\mathbb{R}_+$ is right-continuous and of locally bounded variation, the measure induced by its distribution derivative is denoted df, so that $f(t) = f(0) + \int_{[0,t]} df(s)$ for all $t \geq 0$. Finally, we use the notation $*$ for the convolution operation, we refer to Appendix IV.B for more details.

2 Existence and uniqueness of the Volterra Heston model

We study in this section the existence and uniqueness of the Volterra Heston model given by (1)-(2) allowing for arbitrary curves g_0 as input. When g_0 is given by (3), [AJLP17, Theorem 7.1(i)] provides the existence of a $\mathbb{R}_+^2$-valued weak solution to (1)-(2) under the following mild assumptions on K:

$$\begin{gathered} K \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R}), \text{ and there is } \gamma \in (0,2] \text{ such that } \int_0^h K(t)^2 dt = O(h^\gamma) \text{ and} \\ \int_0^T (K(t+h) - K(t))^2 dt = O(h^\gamma) \text{ for every } T < \infty, \end{gathered} \tag{H_0}$$

$$\begin{gathered} K \text{ is nonnegative, not identically zero, non-increasing and continuous on} \\ (0,\infty), \text{ and its } \textit{resolvent of the first kind } L \text{ is nonnegative and non-increasing}^1 \\ \text{in the sense that } s \to L([s, s+t]) \text{ is non-increasing for all } t \geq 0. \end{gathered} \tag{H_1}$$

We show in Theorem 1 below that weak existence in $\mathbb{R}_+^2$ continue to hold for (1)-(2) for a wider class of admissible input curves g_0. Since S is determined by V, it suffices to study the Volterra square-root equation (2). Theorem 6(ii) in the Appendix guarantees the existence of an unconstrained continuous weak solution V to the following modified equation

$$V_t = g_0(t) + \int_0^t K(t-s) \left(-\lambda V_s ds + \nu \sqrt{V_s^+} dW_s \right), \tag{5}$$

for any locally Hölder continuous function g_0, where $x^+ : x \to \max(0, x)$. Clearly, one needs to impose additional assumptions on g_0 to ensure the nonnegativity of V and drop the positive

[1] We refer to Appendix IV.B for the definition of the *resolvent of the first kind* and some of its properties.

part in (5) so that V solves (2). Hence, weak existence of a nonnegative solution to (2) boils down to finding a set $\mathcal{G}_K$ of admissible input curves g_0 such that any solution V to (5) is nonnegative.

To get a taste of the admissible set $\mathcal{G}_K$, we start by assuming that g_0 and K are continuously differentiable on $[0,\infty)$. In that case, V is a semimartingale such that

$$dV_t = \left(g_0'(t) + (K' * dZ)_t - K(0)\lambda V_t\right) dt + K(0)\nu\sqrt{V_t^+}\, dW_t, \tag{6}$$

where $Z = \int_0^{\cdot}(-\lambda V_s ds + \nu\sqrt{V_s^+} dW_s)$. Relying on Lemma 1 in the Appendix[2], we have

$$K' = (K' * L)(0)K + d(K' * L) * K,$$

so that $K' * dZ$ can be expressed as a functional of (V, g_0) as follows

$$K' * dZ = (K' * L)(0)(V - g_0) + d(K' * L) * (V - g_0). \tag{7}$$

Since $V_0 = g_0(0)$, it is straightforward that $g_0(0)$ should be nonnegative. Now, assume that V hits zero for the first time at $\tau \geq 0$. After plugging (7) in the drift of (6), a first-order Euler scheme leads to the formal approximation

$$V_{\tau+h} \approx \left(g_0'(\tau) - (K' * L)(0) g_0(\tau) - (d(K' * L) * g_0)(\tau) + (d(K' * L) * V)_\tau\right) h,$$

for small $h \geq 0$. Since $K' * L$ is non-decreasing and $V \geq 0$ on $[0,\tau]$, it follows that $(d(K' * L) * V)_\tau \geq 0$ yielding the nonnegativity of $V_{\tau+h}$ if we impose the following additional condition

$$g_0' - (K' * L)(0) g_0 - d(K' * L) * g_0 \geq 0.$$

In the general case, V is not necessarily a semimartingale, and a delicate analysis should be carried on the integral equation (5) instead of the infinitesimal version (6). This suggests that the infinitesimal derivative operator should be replaced by the semigroup operator of right shifts leading to the following condition on g_0

$$\Delta_h g_0 - (\Delta_h K * L)(0) g_0 - d(\Delta_h K * L) * g_0 \geq 0, \quad h \geq 0,^3 \tag{8}$$

and to the following definition of the set $\mathcal{G}_K$ of admissible input curves

$$\mathcal{G}_K = \left\{g_0 \in \mathcal{H}^{\gamma/2} \text{ satisfying (8) and } g_0(0) \geq 0\right\}, \tag{9}$$

where $\mathcal{H}^\alpha = \{g_0 : \mathbb{R}_+ \to \mathbb{R},$ locally Hölder continuous of any order strictly smaller than $\alpha\}$. Recall that γ is the exponent associated with K in (H_0).

The following theorem establishes the existence of a $\mathbb{R}_+^2$-valued weak continuous solution to (1)-(2) on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ for any admissible input curve $g_0 \in \mathcal{G}_K$. Since S is determined by V, the proof follows directly from Theorems 6-7.

[2] Under (H_1) one can show that $K' * L$ is right-continuous, non-decreasing and of locally bounded variation (as in Remark 3 in the Appendix), thus the associated measure $d(K' * L)$ is well defined.

[3] Recall that under (H_1) one can show that $\Delta_h K * L$ is right-continuous and of locally bounded variation (see Remark 3 in the Appendix), thus the associated measure $d(\Delta_h K * L)$ is well defined.

Theorem 1. *Assume that K satisfies (H_0)-(H_1). Then, the stochastic Volterra equation (1)-(2) has a $\mathbb{R}^2_+$-valued continuous weak solution (S,V) for any positive initial condition S_0 and any admissible input curve $g_0 \in \mathscr{G}_K$. Furthermore, the paths of V are locally Hölder continuous of any order strictly smaller than $\gamma/2$ and*

$$\sup_{t\leq T}\mathbb{E}[|V_t|^p] < \infty, \quad p>0, \quad T>0. \tag{10}$$

Example IV.1. *The following classes of functions belong to $\mathscr{G}_K$.*

(i) $g \in \mathscr{H}^{\gamma/2}$ *non-decreasing such that* $g(0)\geq 0$. *Since K is non-increasing and L is nonnegative, we have $0 \leq \Delta_h K * L \leq 1$ for all $h \geq 0$ (see the proof of [AJLP17, Theorem 3.5]) yielding, for all $t,h\geq 0$, that $\Delta_h g(t) - (\Delta_h K * L)(0)g(t) - (d(\Delta_h K * L) * g)(t)$ is equal to*

$$\int_0^t (g(t)-g(t-s))(\Delta_h K * L)(ds) + g(t+h) - g(t) + g(t)(1-(\Delta_h K * L)(t)) \geq 0$$

(ii) $g = V_0 + K * \theta$, *with $V_0 \geq 0$ and $\theta \in \mathbb{L}^2_{loc}(\mathbb{R}_+,\mathbb{R})$ such that $\theta(s)ds + V_0 L(ds)$ is a nonnegative measure. First, $g \in \mathscr{H}^{\gamma/2}$ due to* (H_0) *and the Cauchy-Schwarz inequality*

$$(g(t+h)-g(t))^2 \leq 2\left(\int_0^t (K(s+h)-K(s))^2 ds + \int_0^h K(s)^2 ds\right)\int_0^{t+h}\theta(s)^2 ds.$$

Moreover, $g(0) = V_0 \geq 0$ *and*

$$\Delta_h g - (\Delta_h K * L)(0) g - d(\Delta_h K * L) * g \tag{11}$$

is equal to

$$V_0(1-\Delta_h K * L) + \Delta_h(K*\theta) - (\Delta_h K * L)(0) K*\theta - d(\Delta_h K * L) * K * \theta.$$

(8) *now follows from Lemma 2 with $F = \Delta_h K$, after noticing that* (11) *becomes*

$$\Delta_h(K*(V_0 L + \theta)) - \Delta_h K * (V_0 L + \theta) = \int_\cdot^{\cdot+h} K(\cdot + h - s)(V_0 L(ds) + \theta(s)ds) \geq 0.$$

We now tackle the weak uniqueness of (1)-(2) by characterizing the Fourier-Laplace transform of the process $X = (\log S, V)$. Indeed, when g_0 is of the form (3), X is a two-dimensional affine Volterra process in the sense of [AJLP17, Definition 4.1]. For this particular g_0, [AJLP17, Theorem 7.1(ii)] provides the exponential-affine transform formula

$$\mathbb{E}[\exp(uX_T + (f*X)_T)] = \exp\left(\psi_1(T)\log S_0 + u_2 g_0(T) + \int_0^T F(\psi_1,\psi_2)(s) g_0(T-s)ds\right) \tag{12}$$

for suitable $u \in (C^2)^*$ and $f \in \mathbb{L}^1([0,T],(C^2)^*)$ with $T>0$, where $\psi = (\psi_1,\psi_2)$ solves the following system of Riccati-Volterra equations

$$\psi_1 = u_1 + 1 * f_1, \tag{13}$$

$$\psi_2 = u_2 K + K * F(\psi_1,\psi_2), \tag{14}$$

with

$$F(\psi_1,\psi_2) = f_2 + \frac{1}{2}\left(\psi_1^2 - \psi_1\right) + (\rho\nu\psi_1 - \lambda)\psi_2 + \frac{\nu^2}{2}\psi_2^2. \tag{15}$$

A straightforward adaptation of [AJLP17, Theorems 4.3 and 7.1] shows that the affine transform (12) carries over for any admissible input curve $g_0 \in \mathcal{G}_K$ with the same Riccati equations (13)-(14).

Theorem 2. *Assume that K satisfies (H_0) and that the shifted kernels $\Delta_h K$ satisfy (H_1) for all $h \in [0,1]$. Fix $g_0 \in \mathcal{G}_K, S_0 > 0$ and denote by (S,V) a $\mathbb{R}_+^2$-valued continuous weak solution to* (1)-(2). *For any $u \in (\mathbb{C}^2)^*$ and $f \in \mathbb{L}^1_{loc}(\mathbb{R}_+,(\mathbb{C}^2)^*))$ such that*

$$Re\,\psi_1 \in [0,1],\ Re\,u_2 \leq 0 \text{ and } Re\,f_2 \leq 0, \tag{16}$$

with ψ_1 given by (13), *the Riccati–Volterra equation* (14) *admits a unique global solution $\psi_2 \in \mathbb{L}^2_{loc}(\mathbb{R}_+,\mathbb{C}^*)$. Moreover, the exponential-affine transform* (12) *is satisfied. In particular, weak uniqueness holds for* (1)-(2).

3 Markovian structure

Using the same methodology as in Chapter III, we characterize the Markovian structure of the Volterra Heston model (1)-(2) in terms of the $\mathbb{F}$-adapted infinite-dimensional adjusted forward curve $(g_t)_{t\geq 0}$ given by (4) which is well defined thanks to (10). Furthermore, we prove that the set $\mathcal{G}_K$ is stochastically invariant with respect to $(g_t)_{t\geq 0}$.

Theorem 3. *Under the assumptions of Theorem 1, fix $g_0 \in \mathcal{G}_K$. Denote by (S,V) the unique solution to* (1)-(2) *and by $(g_t)_{t\geq 0}$ the process defined by* (4). *Then, (S^{t_0}, V^{t_0}) satisfies*

$$dS_t^{t_0} = S_t^{t_0}\sqrt{V_t^{t_0}}\,dB_t^{t_0}, \quad S_0^{t_0} = S_{t_0},$$
$$V_t^{t_0} = g_{t_0}(t) + \int_0^t K(t-s)\left(-\lambda V_s^{t_0} ds + \nu\sqrt{V_s^{t_0}}\,dW_s^{t_0}\right),$$

where $(B^{t_0}, W^{t_0}) = (B_{t_0+\cdot} - B_{t_0}, W_{t_0+\cdot} - W_{t_0})$ are two Brownian motions independent of $\mathcal{F}_{t_0}$ such that $d\langle B^{t_0}, W^{t_0}\rangle_t = \rho dt$. Moreover, $\mathcal{G}_K$ is stochastically invariant with respect to $(g_t)_{t\geq 0}$, that is

$$g_t \in \mathcal{G}_K, \quad t \geq 0.$$

Proof. The part for V^{t_0} is immediate after observing that

$$g_{t_0}(t) = g_0(t_0+t) - \int_0^{t_0} K(t+t_0-s)\lambda V_s ds + \int_0^{t_0} K(t+t_0-s)\nu\sqrt{V_s}\,dW_s, \tag{17}$$

for all $t_0, t, h \geq 0$. The part for S^{t_0} is straightforward. We move to proving the claimed invariance. Fix $t_0, t, h \geq 0$ and define $Z = \int_0^{\cdot}(-\lambda V_s ds + \nu\sqrt{V_s}dW_s)$. By Lemma 2 and Remark 3 in the Appendix,

$$\Delta_h K = (\Delta_h K * L)(0)K + d(\Delta_h K * L) * K, \tag{18}$$

so that

$$(\Delta_h K * dZ) = (\Delta_h K * L)(0)(V - g_0) + d(\Delta_h K * L) * (V - g_0).$$

Hence,

$$\begin{aligned}
V_{t+h}^{t_0} &= g_0(t_0+t+h) + (\Delta_h K * dZ)_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&= g_0(t_0+t+h) + (\Delta_h K * L)(0)(V_t^{t_0} - g_0(t_0+t)) \\
&\quad + \left(d(\Delta_h K * L) * (V - g_0)\right)_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&= g_0(t_0+t+h) - (\Delta_h K * L)(0) g_0(t_0+t) - \left(d(\Delta_h K * L) * g_0\right)(t_0+t) \\
&\quad + (\Delta_h K * L)(0) V_t^{t_0} + (d(\Delta_h K * L) * V)_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&\geq (\Delta_h K * L)(0) V_t^{t_0} + (d(\Delta_h K * L) * V)_{t_0+t} - \int_0^h K(h-s)\lambda V_{t+s}^{t_0} ds \\
&\quad + \int_0^h K(h-s)\nu\sqrt{V_{t+s}^{t_0}} dW_{t+s}^{t_0},
\end{aligned}$$

since $g_0 \in \mathcal{G}_K$. We now prove (8). Set $G_h^{t_0} = \Delta_h g_{t_0} - (\Delta_h K * L)(0) g_{t_0} - d(\Delta_h K * L) * g_{t_0}$. The previous inequality combined with (4) yields

$$\begin{aligned}
G_h^{t_0}(t) &= \mathbb{E}\left[V_{t+h}^{t_0} + (\lambda K * V^{t_0})_{t+h} - (\Delta_h K * L)(0)(V_t^{t_0} + (\lambda K * V^{t_0})_t) \,\middle|\, \mathcal{F}_{t_0}\right] \\
&\quad - \mathbb{E}\left[\left(d(\Delta_h K * L) * (V^{t_0} + \lambda K * V^{t_0})\right)_t \,\middle|\, \mathcal{F}_{t_0}\right] \\
&\geq \mathbb{E}\left[(d(\Delta_h K * L) * V)_{t_0+t} - \left(d(\Delta_h K * L) * V^{t_0}\right)_t - \int_0^h K(h-s)\lambda V_{t+s}^{t_0} ds \,\middle|\, \mathcal{F}_{t_0}\right] \\
&\quad + \mathbb{E}\left[(\lambda K * V^{t_0})_{t+h} - \left(((\Delta_h K * L)(0) K + d(\Delta_h K * L) * K) * \lambda V^{t_0}\right)_t \,\middle|\, \mathcal{F}_{t_0}\right].
\end{aligned}$$

Relying on (18), we deduce

$$\begin{aligned}
G_h^{t_0}(t) &\geq \mathbb{E}\left[\int_t^{t_0+t} (d(\Delta_h K * L))(ds) V_{t_0+t-s} - \int_0^h K(h-s)\lambda V_{t+s}^{t_0} ds \,\middle|\, \mathcal{F}_{t_0}\right] \\
&\quad + \mathbb{E}\left[\int_t^{t+h} K(t+h-s)\lambda V_s^{t_0} ds \,\middle|\, \mathcal{F}_{t_0}\right] \\
&= \mathbb{E}\left[\int_t^{t_0+t} (d(\Delta_h K * L))(ds) V_{t_0+t-s} \,\middle|\, \mathcal{F}_{t_0}\right].
\end{aligned}$$

Hence (8) holds for g_{t_0}, since $V \geq 0$ and $d(\Delta_h K * L)$ is a nonnegative measure, see Remark 3. Finally, by adapting the proof of [AJLP17, Lemma 2.4], we can show that for any $p > 1, \epsilon > 0$ and $T > 0$, there exists a positive constant C_1 such that

$$\mathbb{E}\left[|V_{t+h} - V_t|^p\right] \leq C_1 h^{p(\gamma/2-\epsilon)}, \quad t, h \geq 0,\ t+h \leq T + t_0,$$

Relying on (H_0), (4) and Jensen inequality, there exists a positive constant C_2 such that

$$\mathbb{E}\left[|g_{t_0}(t+h)-g_{t_0}(t)|^p\right] \le C_2 h^{p(\gamma/2-\epsilon)}, \quad t,h\ge 0,\ t+h\le T,$$

By Kolmogorov continuity criterion, $g_{t_0}\in\mathcal{H}^{\gamma/2}$ so that $g_{t_0}\in\mathcal{G}_K$ since $g_{t_0}(0)=V_{t_0}\ge 0$. □

Theorem 3 highlights that V is Markovian in the state variable $(g_t)_{t\ge0}$. Indeed, conditional on $\mathcal{F}_t$ for some $t\ge 0$, the shifted Volterra Heston model (S^t,V^t) can be started afresh from (S_t,g_t) with the same dynamics as in (1)-(2). Notice that g_t is again an admissible input curve belonging to $\mathcal{G}_K$. Therefore, applying Theorems 1 and 2 with (S^t,V^t,g_t) yields that the conditional Fourier-Laplace transform of $X=(\log S,V)$ is exponentially affine in $(\log S_t,g_t)$:

$$\mathbb{E}\left[\exp(uX_T+(f*X)_T)\;\middle|\;\mathcal{F}_t\right]=\exp\left((\Delta_{T-t}f*X)_t+\psi_1(T-t)\log S_t+(u_2g_t+F(\psi_1,\psi_2)*g_t)(T-t)\right), \tag{19}$$

for all $t\le T$, where F is given by (15), under the standing assumptions of Theorem 2.

Moreover, it follows from (17) and the fact that $g_\cdot(0)=V$ that the process $(g_t)_{t\ge0}$ solves

$$g_t(x)=\Delta_t g_0(x)+\int_0^t \Delta_{t-s}\left(-\lambda K g_s(0)\right)(x)ds+\int_0^t \Delta_{t-s}\left(K\nu\sqrt{g_s(0)}\right)(x)dW_s. \tag{20}$$

Recalling that $(\Delta_t)_{t\ge0}$ is the semigroup of right shifts, (20) can be seen as a $\mathcal{G}_K$-valued mild solution of the following Heath–Jarrow–Morton-type *stochastic partial differential equation*

$$dg_t(x)=\left(\frac{d}{dx}g_t(x)-\lambda K(x)g_t(0)\right)dt+K(x)\nu\sqrt{g_t(0)}dW_t, \quad g_0\in\mathcal{G}_K. \tag{21}$$

The following proposition provides the characteristic functional of $(g_t)_{t\ge0}$ leading to the strong Markov property of $(g_t)_{t\ge0}$. Define $\langle g,h\rangle=\int_{\mathbb{R}_+}g(x)h(x)dx$, for suitable functions f and g.

Theorem 4. *Under the assumptions of Theorem 2. Let $h\in\mathcal{C}_c^\infty(\mathbb{R}_+)$ and $g_0\in\mathcal{G}_K$. Then,*

$$\mathbb{E}\left[\exp\left(i\langle g_t,h\rangle\right)\right]=\exp\left(\langle H_t,g_0\rangle\right), \quad t\ge 0, \tag{22}$$

where H solves

$$H_t(x)=ih(x-t)\mathbb{1}_{\{x>t\}}+\mathbb{1}_{\{x\le t\}}\left(-\lambda\langle H_{t-x},K\rangle+\frac{\nu^2}{2}\langle H_{t-x},K\rangle^2\right), \quad t,x\ge 0. \tag{23}$$

In particular, weak uniqueness holds for (20) and $(g_t)_{t\ge0}$ is a strong Markov process on $\mathcal{G}_K$.

Proof. Consider $\widetilde{S}_t=1+\int_0^t\widetilde{S}_u\sqrt{V_u}dW_u$, for all $t\ge0$. Then, $(\widetilde{S},V)$ is a Volterra Heston model of the form (1)-(2) with $\rho=1$ and $\widetilde{S}_0=1$. Fix $t\ge0$, $\langle g_t,h\rangle$ is well defined since $x\to g_t(x)$ is continuous. It follows from (17) together with stochastic Fubini theorem, see [Ver12, Theorem 2.2], which is justified by (10), that

$$\begin{aligned}\langle g_t,h\rangle&=\langle g_0(t+\cdot),h\rangle+\left(\frac{\nu}{2}-\lambda\right)\int_0^t\langle K(t-s+\cdot),h\rangle V_s ds+\nu\int_0^t\langle K(t-s+\cdot),h\rangle d(\log\widetilde{S})_s\\&=\langle g_0,h(-t+\cdot)\rangle+\left(\frac{\nu}{2}-\lambda\right)\int_0^t\langle K,h(s-t+\cdot)\rangle V_s ds\\&\quad+\nu\langle K,h\rangle\log\widetilde{S}_t-\nu\int_0^t\langle K,h'(s-t+\cdot)\rangle\log\widetilde{S}_s ds,\end{aligned}$$

where the last identity follows from an integration by parts. Hence, setting

$$
\begin{aligned}
&u_2 = 0, \quad u_1 = \mathrm{i}\nu\langle K, h\rangle, \quad f_1(t) = -\mathrm{i}\nu\langle K, h'(-t+\cdot)\rangle,\\
&\psi_1(t) = u_1 + (1 * f_1)(t) = \mathrm{i}\ \nu\langle K(t+\cdot), h\rangle,\\
&f_2(t) = \mathrm{i}\ (\frac{\nu}{2} - \lambda)\langle K(t+\cdot), h\rangle, \quad \psi_2 = K * F(\psi_1, \psi_2),
\end{aligned}
$$

with F as in (15), the characteristic functional follows from Theorem 2

$$
\mathbb{E}\left[\exp\left(\mathrm{i}\langle g_t, h\rangle\right)\right] = e^{\mathrm{i}\langle h(-t+\cdot), g_0\rangle}\mathbb{E}\left[\exp\left(u_1 \log \widetilde{S}_t + (f_1 * \log \widetilde{S})_t + (f_2 * V)_t\right)\right] = \exp\left(\langle H_t, g_0\rangle\right)
$$

where

$$
H_t(x) = h(x-t)\mathbf{1}_{\{x>t\}} + \mathbf{1}_{\{0\le x\le t\}} F(\psi_1, \psi_2)(t-x), \quad x \ge 0,
$$

and (15) reads

$$
\begin{aligned}
F(\psi_1, \psi_2)(t) = &-\lambda\langle K(t+\cdot), h\rangle + \frac{\nu^2}{2}\langle K(t+\cdot), h\rangle^2\\
&+ (\nu^2\langle K(t+\cdot), h\rangle - \lambda)\psi_2(t) + \frac{\nu^2}{2}\psi_2(t)^2.
\end{aligned}
\tag{24}
$$

Now observe that

$$
\langle H_t, K\rangle = \langle h(-t+\cdot), K\rangle + \int_0^t F(\psi_1, \psi_2)(t-x)K(x)dx = \langle h, K(t+\cdot)\rangle + \psi_2(t).
$$

Hence, after plugging $\psi_2(t) = \langle H_t, K\rangle - \langle h, K(t+\cdot)\rangle$ back in (24) we get that

$$
F(\psi_1, \psi_2)(t) = -\lambda\langle H_t, K\rangle + \frac{\nu^2}{2}\langle H_t, K\rangle^2,
$$

yielding (23). Weak uniqueness now follows by standard arguments. In fact, thanks to (10) and stochastic Fubini theorem, $(g_t)_{t\ge 0}$ solves (21) in the weak sense, that is

$$
\langle g_t, h\rangle = \langle g_0, h\rangle + \int_0^t \left(\langle g_s, -h'\rangle - \lambda\langle K, h\rangle g_s(0)\right) ds + \int_0^t \nu\langle K, h\rangle \sqrt{g_s(0)} dW_s, \quad h \in \mathscr{C}_c^\infty(\mathbb{R}).
$$

Therefore, combined with Theorem 3, $(g_t)_{t\ge 0}$ solves a martingale problem on $\mathscr{G}_K$. In addition, (22) yields uniqueness of the one-dimensional distributions which is enough to get weak uniqueness for (20) and the strong Markov property by [EK86, Theorem 4.4.2]. □

We notice that (22)-(23) agree with [GKR18, Proposition 4.5] when $\lambda = 0$. Moreover, one can lift (23) to a non-linear partial differential equation in duality with (21). Indeed, define the measure-valued function $\bar{H} : t \to \bar{H}_t(dx) = H_t(x)\mathbf{1}_{\{x\ge 0\}}dx$. Then, it follows from (23) that

$$
\bar{H}_t(dx) = \mathrm{i}\ h(x-t)\mathbf{1}_{\{x>t\}}dx + \int_0^t \delta_0(dx-(t-s))(-\lambda\langle \bar{H}_s, K\rangle + \frac{\nu^2}{2}\langle \bar{H}_s, K\rangle^2)ds
$$

which can be seen as the mild formulation of the following partial differential equation

$$d\bar{H}_t(dx) = (-\frac{d}{dx}\bar{H}_t(dx) + \delta_0(dx)(-\lambda\langle\bar{H}_t, K\rangle + \frac{\nu^2}{2}\langle\bar{H}_t, K\rangle^2))dt, \quad \bar{H}_0(dx) = \mathrm{i}h(x)\mathbf{1}_{\{x>t\}}dx. \tag{25}$$

We refer to [CT18] for similar results in the discontinuous setting. The previous results highlight not only the correspondence between stochastic Volterra equations of the form (2) and stochastic partial differential equations (21) but also between their dual objects, that is the Riccati-Volterra equation (14) and the non-linear partial differential equation (25). One can establish a correspondence between (2) and other related stochastic partial differential equations which, unlike $(g_t)_{t\geq 0}$, do not necessarily have a financial interpretation but for which the dual object satisfies a nicer non-linear partial differential equation than (25), see [MS15].

4 Application: square-root process in Banach space

As an application of Theorems 1, 2, 3, we obtain conditions for weak existence and uniqueness of the following (possibly) infinite-dimensional system of stochastic differential equations

$$dU_t(x) = \left(-xU_t(x) - \lambda\int_0^\infty U_t(z)\mu(dz)\right)dt + \nu\sqrt{\int_0^\infty U_t(z)\mu(dz)}dW_t, \quad x \in \operatorname{supp}(\mu), \tag{26}$$

for a fixed positive measure of locally bounded variation μ[4]. This is achieved by linking (26) to a stochastic Volterra equation of the form (2) with the following kernel

$$K(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t > 0. \tag{27}$$

We will assume that μ is a positive measure of locally bounded variation such that

$$\int_0^\infty (1 \wedge (xh)^{-1/2})\mu(dx) \leq Ch^{(\gamma-1)/2}, \quad \int_0^\infty x^{-1/2}(1 \wedge (xh))\mu(dx) \leq Ch^{\gamma/2}; \quad h > 0, \tag{H_2}$$

for some $\gamma \in (0, 2]$ and positive constant C. The reader may check that in that case K satisfies (H_0). Furthermore, [GLS90, Theorem 5.5.4] guarantees the existence of the resolvent of the first kind L of K and that (H_1) is satisfied for the shifted kernels $\Delta_h K$ for any $h \in [0, 1]$. Hence, K satisfies assumptions of Theorems 1 and 2.

By a solution U to (26) we mean a family of continuous processes $(U(x))_{x\in\operatorname{supp}(\mu)}$ such that $x \to U_t(x) \in \mathbb{L}^1(\mu)$ for any $t \geq 0$, $(\int_0^\infty U_t(x)\mu(dx))_{t\geq 0}$ is a continuous process and (26) holds a.s. on some filtered probability space. If such solution exists, we set $V = \int_0^\infty U_\cdot(x)\mu(dx)$ and $g_0 = \int_0^\infty U_0(x)e^{-x(\cdot)}dx$. Thanks to (H_2), the stochastic Fubini theorem yields for each $t \geq 0$

$$V_t = g_0(t) + \int_0^t K(t-s)(-\lambda V_s ds + \nu\sqrt{V_s}dW_s). \tag{28}$$

[4]We use the notation $\operatorname{supp}(\mu)$ to denote the support of a measure μ, that is the set of all points for which every open neighborhood has a positive measure. Here we assume that the support is in $\mathbb{R}_+$.

The processes above being continuous, the equality holds in terms of processes. Thus, provided that g_0 belongs to $\mathscr{G}_K$, Theorem 2 leads to the weak uniqueness of (26) because for each $x \in \mathrm{supp}(\mu)$,

$$U_t(x) = e^{-xt}U_0(x) + \int_0^t e^{-x(t-s)}(-\lambda V_s ds + \nu\sqrt{V_s}dW_s), \quad t \geq 0. \tag{29}$$

On the other hand, if we assume that $g_0 = \int_0^\infty U_0(x)e^{-x(\cdot)}\mu(dx) \in \mathscr{G}_K$ for some initial family of points $(U_0(x))_{x\in\mathrm{supp}(\mu)} \in \mathbb{L}^1(\mu)$, there exists a continuous solution V for (28) by Theorem 1. In that case, we define for each $x \in \mathrm{supp}(\mu)$, the continuous process $U(x)$ as in (29). Thanks to (H_2) and (10), another application of the stochastic Fubini theorem combined with the fact that V satisfies (28) yields that, for each $t \geq 0$, $(U_t(x))_{x\in\mathrm{supp}(\mu)} \in \mathbb{L}^1(\mu)$ and

$$V_t = \int_0^\infty U_t(x)\mu(dx). \tag{30}$$

Moreover, by an integration by parts, we get for each $x \in \mathrm{supp}(\mu)$,

$$U_t(x) = e^{-xt}U_0(x) + Z_t e^{-xt} + \int_0^t xe^{-x(t-s)}(Z_s - Z_t)ds,$$

with $Z = \int_0^\cdot(-\lambda V_s ds + \nu\sqrt{V_s}dB_s)$. We know that for fixed $T > 0$, $\eta \in (0, 1/2)$ and for almost any $\omega \in \Omega$ there exists a positive constant $C_T(\omega)$ such that $|Z_s - Z_t| \leq C_T(\omega)|t-s|^\eta$ for all $t, s \in [0, T]$. Hence for any $t \in [0, T]$ and $x \in \mathrm{supp}(\mu)$

$$|U_t(x)| \leq |U_0(x)| + C_T(\omega)e^{-xt}t^\eta + C_T(\omega)x\int_0^t e^{-xs}s^\eta ds = |U_0(x)| + C_T(\omega)\eta\int_0^t e^{-xs}s^{\eta-1}ds.$$

Then,

$$\sup_{t\in[0,T]}|U_t(x)| \leq |U_0(x)| + C_T(\omega)\eta\int_0^T e^{-xs}s^\eta ds \in \mathbb{L}^1(\mu).$$

Therefore by dominated convergence theorem, the process $(\int_0^\infty U_t(x)\mu(dx))_{t\geq 0}$ is continuous. In particular, (30) holds in terms of processes and it follows from (29) that U is a solution of (26).

This leads to the weak existence and uniqueness of (26) if the initial family of points $(U_0(x))_{x\in\mathrm{supp}(\mu)}$ belongs to the following space $\mathscr{D}_\mu$ defined by

$$\mathscr{D}_\mu = \{(u_x)_{x\in\mathrm{supp}(\mu)} \in \mathbb{L}^1(\mathrm{supp}(\mu)); \quad \int_0^\infty u_x e^{-x(\cdot)}\mu(dx) \in \mathscr{G}_K\}, \tag{31}$$

with K given by (27). Notice that for fixed $t_0 \geq 0$ and for any $t \geq 0$ and $x \in \mathrm{supp}(\mu)$,

$$U_{t+t_0}(x) = U_{t_0}(x)e^{-xt} + \int_0^t e^{-x(t-s)}\left(-\lambda\int_0^\infty U_{s+t_0}(z)\mu(dz) + \nu\sqrt{\int_0^\infty U_{s+t_0}(z)\mu(dz)}dW_{s+t_0}\right)$$

and then by stochastic Fubini theorem

$$\int_0^\infty U_{t+t_0}(y)\mu(dy) = g_{t_0}(t) + \int_0^t K(t-s)\left(-\lambda\int_0^\infty U_{s+t_0}(z)\mu(dz) + \nu\sqrt{\int_0^\infty U_{s+t_0}(z)\mu(dz)}dW_{s+t_0}\right),$$

with $g_{t_0}(t) = \int_0^\infty U_{t_0}(y)e^{-yt}\mu(dy)$. Thanks to Theorem 3, we deduce that $g_{t_0} \in \mathscr{G}_K$ and therefore $(U_{t_0}(x))_{x\in \mathrm{supp}(\mu)}$ belongs to $\mathscr{D}_\mu$. As a conclusion, the space $\mathscr{D}_\mu$ is stochastically invariant with respect to the family of processes $(U(x))_{x\in\mathrm{supp}(\mu)}$.

Theorem 5. *Fix μ a positive measure of locally bounded variation satisfying* (H_2).*There exists a unique weak solution U of* (26) *for each initial family of points* $(U_0(x))_{x\in\mathrm{supp}(\mu)} \in \mathscr{D}_\mu$. *Furthermore for any $t \geq 0$,* $(U_t(x))_{x\in\mathrm{supp}(\mu)} \in \mathscr{D}_\mu$.

	$K(t)$	Parameter restrictions	$\mu(d\gamma)$
Fractional	$c\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$\alpha \in (1/2, 1)$	$c\frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$
Gamma	$ce^{-\lambda t}\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$\lambda \geq 0, \alpha \in (1/2,1)$	$c\frac{(x-\lambda)^{-\alpha}\mathbf{1}_{(\lambda,\infty)}(x)}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$
Exponential sum	$\sum_{i=1}^n c_i e^{-\gamma_i t}$	$c_i, \gamma_i \geq 0$	$\sum_{i=1}^n c_i \delta_{\gamma_i}(dx)$

Table IV.1 – Some measures μ satisfying (H_2) with their associated kernels K. Here $c \geq 0$.

Remark 1 (Representation of V in terms of U). *In a similar fashion one can establish the existence and uniqueness of the following time-inhomogeneous version of* (26)

$$dU_t(x) = \left(-xU_t(x) - \lambda\left(g_0(t) + \langle 1, U_t\rangle_\mu\right)\right)dt + \nu\sqrt{g_0(t) + \langle 1, U_t\rangle_\mu}dW_t, \quad x \in \mathrm{supp}(\mu), \tag{32}$$

whenever

$$g_0 = \widetilde{g_0} + \int_0^\infty e^{-x(\cdot)}U_0(x)\mu(dx) \in \mathscr{G}_K,$$

with $\widetilde{g_0} : \mathbb{R}_+ \to \mathbb{R}$. In this case,

$$\widetilde{g_0}(t+\cdot) + \int_0^\infty e^{-x(\cdot)}U_t(x)\mu(dx) \in \mathscr{G}_K, \quad t \geq 0.$$

In particular, for $U_0 \equiv 0$, $g_0 = \widetilde{g_0} \in \mathscr{G}_K$ and K as in (27)*, the solution V to the stochastic Volterra equation* (2) *and the forward process $(g_t)_{t\geq 0}$ admit the following representations*

$$V_t = g_0(t) + \langle 1, U_t\rangle_\mu, \quad g_{t_0}(t) = g_0(t_0+t) + \langle e^{-t(\cdot)}, U_{t_0}\rangle_\mu, \quad t, t_0 \geq 0, \tag{33}$$

where we used the notation $\langle f, g\rangle_\mu = \int_0^t f(x)g(x)\mu(dx)$. These results are in the spirit of [CC98, HS15].

When μ has finite support, (32) is a finite dimensional diffusion with an affine structure in the sense of [DFS03]. This underlying structure carries over to the case of infinite support and is the reason behind the tractability of the Volterra Heston model.

Remark 2 (Affine structure of $(\log S, V)$ in terms of U). *Let the notations and assumptions of Remark 1 be in force. Relying on the existence and uniqueness of the Riccati-Volterra equation* (14) *one can establish the existence and uniqueness of a differentiable (in time) solution χ_2 to the following (possibly) infinite-dimensional system of Riccati ordinary differential equations*

$$\partial_t \chi_2(t,x) = -x\chi_2(t,x) + F\left(\psi_1(t), \langle \chi_2(t,\cdot), 1\rangle_\mu\right), \quad \chi_2(0,x) = u_2, \quad x \in \operatorname{supp}\mu,\ t \geq 0, \tag{34}$$

such that $\chi_2(t,\cdot) \in \mathbb{L}^1(\mu)$, for all $t \geq 0$ and $t \to \langle \chi_2(t,\cdot), 1\rangle_\mu \in \mathbb{L}^2_{loc}(\mathbb{R}_+)$ with ψ_1 given by (13) *and F by* (15). *Moreover, the unique global solution $\psi_2 \in \mathbb{L}^2_{loc}(\mathbb{R}_+, \mathbb{C}^*)$ to the Riccati–Volterra equation* (14) *admits the following representation*

$$\psi_2 = \int_0^\infty \chi_2(\cdot, x)\mu(dx),$$

where χ_2 is the unique solution to (34). *In particular, combining the equality above with* (19) *and the representation of $(g_t)_{t\geq 0}$ in* (33) *leads to the exponentially-affine functional*

$$\mathbb{E}\left[\exp\left(uX_T + (f*X)_T\right) \,\middle|\, \mathscr{F}_t\right] = \exp\left(\phi(t,T) + \psi_1(T-t)\log S_t + \langle \chi_2(T-t,\cdot), U_t\rangle_\mu\right)$$

for all $t \leq T$ where $\phi(t,T) = (\Delta_{T-t} f * X)_t + (u_2 \Delta_t g_0 + F(\psi_1,\psi_2) * \Delta_t g_0)(T-t)$, (u,f) as in (16) *and U solves* (32).

The representations of this section lead to a generic approximation of the Volterra Heston model by finite-dimensional affine diffusions, see Chapter VIII for the rigorous treatment of these approximations.

IV.A Existence results for stochastic Volterra equations

In this section, we consider the following d-dimensional stochastic Volterra equation

$$X_t = g(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{35}$$

where $K \in L^2_{\text{loc}}(\mathbb{R}, \mathbb{R}^{d\times d})$, W is a m-dimensional Brownian motion, $g : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous with linear growth. By adapting the proofs of [AJLP17, Appendix A] (there, g is constant), we obtain the following existence results. Notice how the domain $\mathscr{G}_K$ defined in (9) enters in the construction of constrained solutions in Theorem 7 below.

Theorem 6. *Under (H_0), assume that $g \in \mathscr{H}^{\gamma/2}$.*

(i) If b and σ are Lipschitz continuous, (35) *admits a unique continuous strong solution X.*

(ii) If b and σ are continuous with linear growth and K admits a resolvent of the first kind L, then (35) *admits a continuous weak solution X.*

In both cases, X is locally Hölder continuous of any order strictly smaller than $\gamma/2$ and

$$\sup_{t\leq T} \mathbb{E}[|X_t|^p] < \infty, \quad p > 0, \quad T > 0. \tag{36}$$

Theorem 7. *Assume that $d = m = 1$ and that the scalar kernel K satisfies (H_0)-(H_1). Assume also that b and σ are continuous with linear growth such that $b(0) \geq 0$ and $\sigma(0) = 0$. Then* (35) *admits a nonnegative continuous weak solution for any $g \in \mathcal{G}_K$.*

Proof. Theorem 6(ii) yields the existence of an unconstrained continuous weak solution X to the following modified equation $X_t = g(t) + \int_0^t K(t-s)b(X_s^+)ds + \int_0^t K(t-s)\sigma(X_s^+)dW_s$. As in the proof of of [AJLP17, Theorem 3.5], it suffices to prove the nonnegativity of X under the stronger condition, that, for some fixed $n \in \mathbb{N}$,

$$x \leq n^{-1} \text{ implies } b(x) \geq 0 \text{ and } \sigma(x) = 0. \tag{37}$$

Set $Z = \int(b(X)dt + \sigma(X)dW)$ and $\tau = \inf\{t \geq 0 \colon X_t < 0\}$. Since $g(0) \geq 0$, $\tau \geq 0$. On $\{\tau < \infty\}$,

$$X_{\tau+h} = g(\tau+h) + (K * dZ)_{\tau+h} = g(\tau+h) + (\Delta_h K * dZ)_\tau + \int_0^h K(h-s)dZ_{\tau+s}, \quad h \geq 0. \tag{38}$$

Using Lemma 2 and Remark 3 below, together with the fact that $X \geq 0$ on $[0,\tau]$,

$$\begin{aligned} g(\tau+h) + (\Delta_h K * dZ)_\tau &= g(\tau+h) + (\Delta_h K * L)(0)(X-g)(\tau) \\ &\quad + (d(\Delta_h K * L) * X)_\tau - (d(\Delta_h K * L) * g)(\tau) \\ &\geq g(\tau+h) - (d(\Delta_h K * L) * g)(\tau) - (\Delta_h K * L)(0)g(\tau), \end{aligned}$$

which is nonnegative. In view of (38) it follows that

$$X_{\tau+h} \geq \int_0^h K(h-s)\,(b(X_{\tau+s})ds + \sigma(X_{\tau+s})dW_{\tau+s}) \tag{39}$$

on $\{\tau < \infty\}$ for all $h \geq 0$. Now, on $\{\tau < \infty\}$, $X_\tau = 0$ and $X_{\tau+h} < 0$ for arbitrarily small h. On the other hand, by continuity there is some $\varepsilon > 0$ such that $X_{\tau+h} \leq n^{-1}$ for all $h \in [0,\varepsilon)$. Thus (37) and (39) yield $X_{\tau+h} \geq 0$ for all $h \in [0,\varepsilon)$. This shows that $\tau = \infty$, ending the proof. $\square$

IV.B Reminder on stochastic convolutions and resolvents

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 by

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

whenever these expressions are well-defined. If F is a function on $\mathbb{R}_+$, we write $K * F = K * (Fdt)$. We can show that $L * F$ is almost everywhere well-defined and belongs to $\mathbb{L}^p_{\text{loc}}(\mathbb{R}_+)$, whenever $F \in \mathbb{L}^p_{\text{loc}}(\mathbb{R}_+)$. Moreover, $(F * G) * L = F * (G * L)$ *a.e.*, whenever $F, G \in \mathbb{L}^1_{loc}(\mathbb{R}_+)$, see [GLS90, Theorem 3.6.1 and Corollary 3.6.2] for further details.

For any continuous semimartingale $M = \int_0^\cdot b_s ds + \int_0^\cdot a_s dB_s$ the convolution $(K * dM)_t = \int_0^t K(t-s)dM_s$ is well-defined as an Itô integral for every $t \geq 0$ such that $\int_0^t |K(t-s)||b_s|ds + \int_0^t |K(t-s)|^2|a_s|^2 ds < \infty$. By stochastic Fubini Theorem, see [AJLP17, Lemma 2.1], we have

$(L * (K * dM)) = ((L * K) * dM)$, $a.s.$ whenever $K \in \mathbb{L}^2_{loc}(\mathbb{R}_+, \mathbb{R})$ and a, b are locally bounded *a.s.*

We define the *resolvent of the first kind* of a $d \times d$-matrix valued kernel K, as the $\mathbb{R}^{d\times d}$-valued measure L on $\mathbb{R}_+$ of locally bounded variation such that $K * L = L * K \equiv \mathrm{id}$, where id stands for the identity matrix, see [GLS90, Definition 5.5.1]. The *resolvent of the first kind* does not always exist. The following results are shown in [AJLP17, Lemma 2.6].

Lemma 1. *Let $K \in \mathbb{L}^2_{loc}(\mathbb{R}_+)$ and $Z = \int_0^\cdot b_s ds + \int_0^\cdot \sigma_s dW_s$ a continuous semimartingale with b and σ locally bounded. Assume that X and $K * dZ$ are continuous processes and that K admits a resolvent of the first kind L. Then $X = K * dZ$ if and only if $L * X = Z$.*

Lemma 2. *Assume that $K \in \mathbb{L}^1_{loc}(\mathbb{R}_+)$ admits a resolvent of the first kind L. For any $F \in \mathbb{L}^1_{loc}(\mathbb{R}_+)$ such that $F * L$ is right-continuous and of locally bounded variation one has*

$$F = (F * L)(0)K + d(F * L) * K.$$

Remark 3. *The previous lemma will be used with $F = \Delta_h K$, for a fixed $h \geq 0$. If K is continuous on $(0, \infty)$, then $\Delta_h K * L$ is right-continuous. Moreover, if K is nonnegative and L is non-increasing in the sense that $s \to L([s, s+t])$ is non-increasing for all $t \geq 0$, then $\Delta_h K * L$ is non-decreasing since $\Delta_h K * L = 1 - \int_{(0,h]} K(h-s)L(\cdot + ds)$, $t \geq 0$.*

Part IV

The rough Heston model in practice

CHAPTER V

Roughening Heston

Abstract

Rough volatility models are known to fit the volatility surface remarkably well with very few parameters. On the other hand, the classical Heston model is highly tractable allowing for fast calibration. We present here the rough Heston model which offers the best of both worlds. Even better, we find that we can accurately approximate rough Heston model values by scaling the volatility of volatility parameter of the classical Heston model.

1 Introduction

Rough volatility models have succeeded in capturing the imagination of both practitioners and academics with their remarkable ability to consistently model both historical and implied volatilities with very few parameters. However, even with the introduction of the efficient hybrid BSS scheme of [BLP17], practical implementation has proved to be difficult. In the recent advances in Chapters I, II and III, the authors show how a natural rough generalization of the Heston model emerges as the macroscopic limit of a simple high frequency trading model which reflects the persistence of order flow, the high degree of endogeneity of the market, and liquidity asymmetry between bid and ask sides of the limit order book. In addition, they derive the characteristic function of the log-price as well as hedging strategies in this model.

In this note, we present the rough Heston model and explain how to use it in practice. As in the rough Bergomi model of [BFG16], the forward variance curve $\xi_t(u) = \mathbb{E}[V_u|\mathcal{F}_t]$, where V_u is the spot variance at time u, is a state variable so that the model can be made to match at-the-money volatilities exactly. We are left with only three parameters to calibrate, defined in Section 2: the Hurst exponent H, the volatility of volatility ν and the correlation ρ between spot moves and volatility moves.

2 The rough Heston model

The rough Heston model of Chapter II for a one-dimensional asset price S takes the form

$$\frac{dS_t}{S_t} = \sqrt{V_t}\{\rho\, dB_t + \sqrt{1-\rho^2}\, dB_t^{\perp}\}$$

with

$$V_u = V_t + \frac{\lambda}{\Gamma(H+\frac{1}{2})}\int_t^u \frac{\theta^t(s) - V_s}{(u-s)^{\frac{1}{2}-H}}\, ds + \frac{\nu}{\Gamma(H+\frac{1}{2})}\int_t^u \frac{\sqrt{V_s}}{(u-s)^{\frac{1}{2}-H}}\, dB_s, \quad u \geq t, \tag{1}$$

where $H \in (0, 1/2)$ is the Hurst exponent, ν is the volatility of volatility, $\rho \in [-1, 1]$ is the correlation between spot and volatility moves, $\lambda \geq 0$ and Γ denotes the Gamma function. The mean reversion level parameter $\theta^t(\cdot)$ is allowed to be an $\mathscr{F}_t$-measurable function which makes the model time consistent as explained in Chapter III. It is straightforward to verify that Equation (1) gives back the classical Heston model with time-dependent mean reversion level in the limit $H \to 1/2$. From Chapter III, volatility sample paths have Hölder regularity $H - \varepsilon$ for any $\varepsilon > 0$, hence the name *rough Heston* for this model.

It is also shown in Chapter III that $\lambda\theta^t(\cdot)$ can be directly inferred from the forward variance curve $(\xi_t(u))_{u\geq t}$ observed at time t. By doing so, the model may be rewritten in the asymptotic setting $\lambda \to 0$ in forward variance form as

$$\frac{dS_t}{S_t} = \sqrt{V_t}\{\rho\, dB_t + \sqrt{1-\rho^2}\, dB_t^{\perp}\}$$

with

$$V_u = \xi_t(u) + \frac{\nu}{\Gamma(H+\frac{1}{2})}\int_t^u \frac{\sqrt{V_s}}{(u-s)^{\frac{1}{2}-H}}\, dB_s, \quad u \geq t. \tag{2}$$

Remark 1. *Recall that the forward variance curve may in principle be obtained from the variance swap curve by differentiation. More practically, assuming continuous sample paths, it is well-known that the fair value of variance swaps can be obtained from an infinite log-strip of out-of-the-money options, see for example [Gat11].*

3 Pricing and hedging

Just as in the classical case, we can compute in quasi-closed form the characteristic function of the log-price of the stock in the rough Heston model. This makes the model highly tractable and easy to calibrate.

We define the fractional integral of order $r \in (0, 1]$ of a function f as

$$I^r f(t) = \frac{1}{\Gamma(r)}\int_0^t (t-s)^{r-1} f(s) ds,$$

whenever the integral exists, and its fractional derivative of order $r \in [0,1)$ as

$$D^r f(t) = \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s) ds,$$

whenever it exists. From Chapter III, the rough Heston model is Markovian in $X_t = \log(S_t)$ and the forward variance curve $(\xi_t(u))_{u \geq 0}$, in the same spirit as the Bergomi model [Ber05]. In forward-variance form, the characteristic function of the terminal log-spot X_T conditional on the time t initial state (X_t, ξ_t) is given by

$$\phi_t(T,a) = \mathbb{E}_{X_t,\xi_t}\left[\exp\left\{i\, a X_T\right\}\right] = \exp\left\{i\, a X_t + \int_t^T D^\alpha h(a, T-u)\, \xi_t(u)\, du\right\}, \tag{3}$$

where $\alpha = H + \frac{1}{2}$ and $h(a,.)$ is the unique continuous solution of the fractional Riccati equation

$$D^\alpha h(a,t) = -\frac{1}{2} a(a+i) + i \rho \nu a h(a,t) + \frac{1}{2} \nu^2 h^2(a,t); \quad I^{1-\alpha} h(a,0) = 0. \tag{4}$$

Equation (4) is a rough version of the Riccati equation arising in the classical Heston model (with zero mean reversion). Indeed the only difference is that the time derivative is replaced by a fractional derivative. In contrast to the classical Heston case, there is no explicit solution to (4). This equation can be solved efficiently using numerical methods for fractional ordinary differential equations. We present one such method, the Adams scheme, in Appendix V.A, see [CGP18] and Chapter VIII for newly developed alternative numerical methods. Moreover, as we explain in Section 5, the true solution may also be accurately approximated in closed-form by a scaled version of the classical Heston solution. Then European option prices may be obtained from the characteristic function using standard Fourier techniques, see for example [Gat11].

With the characteristic function now in the Markovian form (3), hedging European options becomes obvious. Let $C_t(T) = \mathbb{E}\left[f(X_T)\middle|\mathscr{F}_t\right]$. Then, a European option with payoff $f(X_T)$ can be perfectly replicated. As of time t, the hedge portfolio has $\partial_{S_t} C_t(T)$ of stock and $\partial_{\xi_t} C_t(T)$ of the forward variance curve $\xi_t(s)$ for each $s \in (t,T]$, where ∂_{ξ_t} represents the Fréchet derivative, roughly speaking the portfolio corresponding to bumping each of the forward variances $\xi_t(s)$, see Chapter III. From the above expressions, it is clear that perfect replication is in theory only. In practice, as with interest rates, one holds a finite number of variance contracts. Note however, that the rough Heston model has only one volatility factor. Thus one can hedge with only one European option, as in the classical Heston case, provided the value of the option component of the hedge portfolio coincides with the theoretical value of the forward variance component.

4 Calibration of the rough Heston model

In this section we present SPX volatility surface calibration results for two dates: August 14, 2013, to compare with the rough Bergomi calibration given for that day in [BFG16], and May

19, 2017, to show that the model continues to fit the market very well.

Note that the only parameters that we calibrate are the Hurst parameter H, the volatility of volatility ν and the correlation ρ. The forward variance curve, being a state variable in the model, is fixed to match at-the-money volatilities.

4.1 SPX calibration on August 14, 2013

From Bloomberg, on August 14, 2013, there were 19 expirations from 1 day to over 2.5 years, for a total of 1,809 options quoted. After eliminating options for which the bid price (of either the put or the call) was zero, we are left with 1,290 strike-expiration pairs.

For each such strike-expiration pair, we compute bid and ask implied volatilities $\sigma_i^{\pm} := \sigma_{\mathrm{BS}}^{\pm}(k_i, T_i)$, where k_i denotes the log-strike. Given model parameters $\{H, \nu, \rho\}$, we can obtain model implied volatilities $\sigma_i^M(H, \nu, \rho)$. We calibrate the parameters by minimizing[1]

$$\sum_i [\sigma_i^M(H,\nu,\rho) - \sigma_i^+]^2 + [\sigma_i^M(H,\nu,\rho) - \sigma_i^-]^2,$$

subject to the constraints

$$H \in (0, 1/2], \quad \nu \geq 0, \quad \rho \in [-1, 1].$$

We obtain the following optimal parameters:

$$H = 0.1216; \quad \nu = 0.2910; \quad \rho = -0.6714.$$

Figure V.1 shows a remarkable fit to the SPX volatility surface, which is as good as with the rough Bergomi model in [BFG16].

4.2 SPX calibration on May 19, 2017

From Bloomberg, on May 19, 2017, there were 27 expirations from 1 day to over 2.5 years, for a total of 2,743 options quoted. By applying the same calibration procedure, we obtain the following optimal parameters:

$$H = 0.0474; \quad \nu = 0.4061; \quad \rho = -0.6710.$$

On this particular day, the calibrated value of H is closer to zero and so corresponds to very rough volatility.

[1] Alternatively, the parameters of the rough Heston model may be efficiently calibrated to the term structure of leverage swaps using a closed-form formula, see [AGR17].

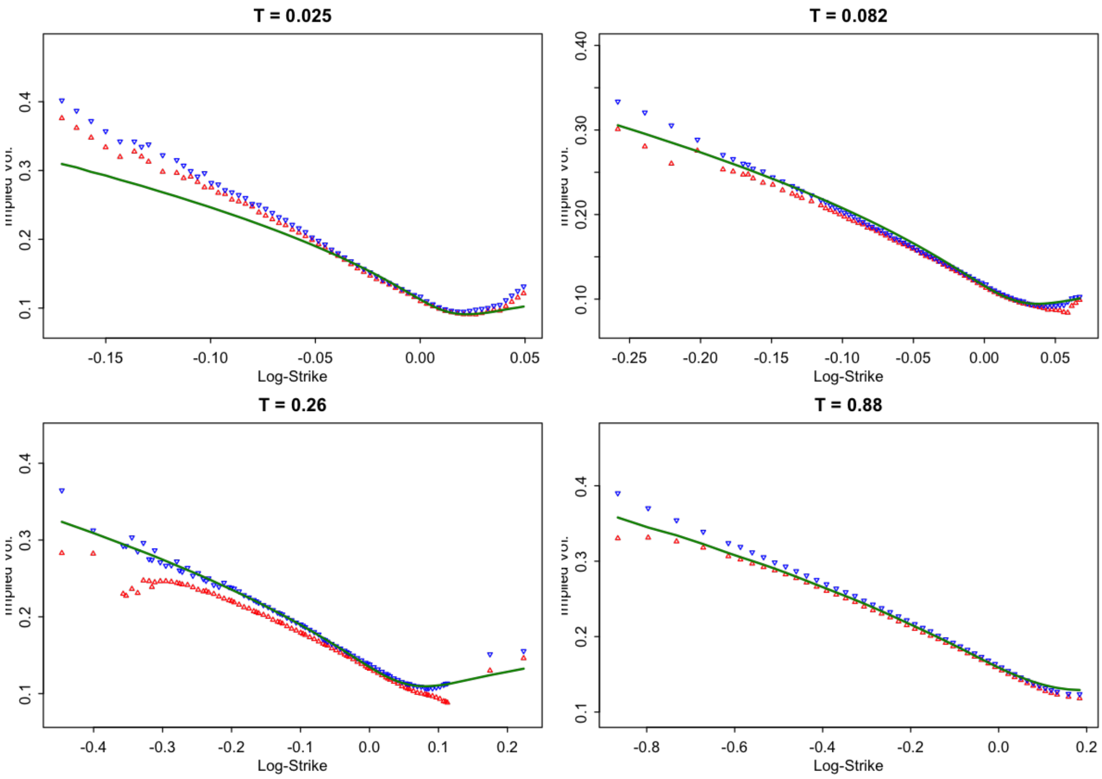

Figure V.1 – Representative SPX volatility smiles as of August 14, 2013, with time to expiration in years. Red and blue points represent bid and ask SPX implied volatilities; green smiles are from the rough Heston model calibrated using the Adams scheme.

4.3 Consistency with historical data

In addition, we have that the calibrated Hurst parameter on August 14, 2013 (see Section 4.1) is consistent with the one computed from historical data. Indeed, in [GJR18], the authors show that the behavior of historical log-volatility of the SPX index is close to that of a fractional Brownian motion with small Hurst parameter of order 0.1. As explained in detail in [GJR18], estimating the moment of order q of a log-volatility increment over a time interval of length Δ by

$$m(\Delta, q) = \frac{1}{N}\sum_{k=1}^{N} |\log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta})|^q,$$

where the $(\sigma_{k\Delta})_{0\le k\le N}$ are historical measurements of volatility, we obtain a strong linear relationship between $\log(m(\Delta, q))$ and $\log(\Delta)$. In summary, we find

$$\mathbb{E}[|\sigma_\Delta - \sigma_0|^q] \approx K_q \Delta^{qH}; \quad H \approx 0.14,$$

which is in theory what we have if the log-volatility is a fractional Brownian motion with Hurst parameter $H = 0.14$.

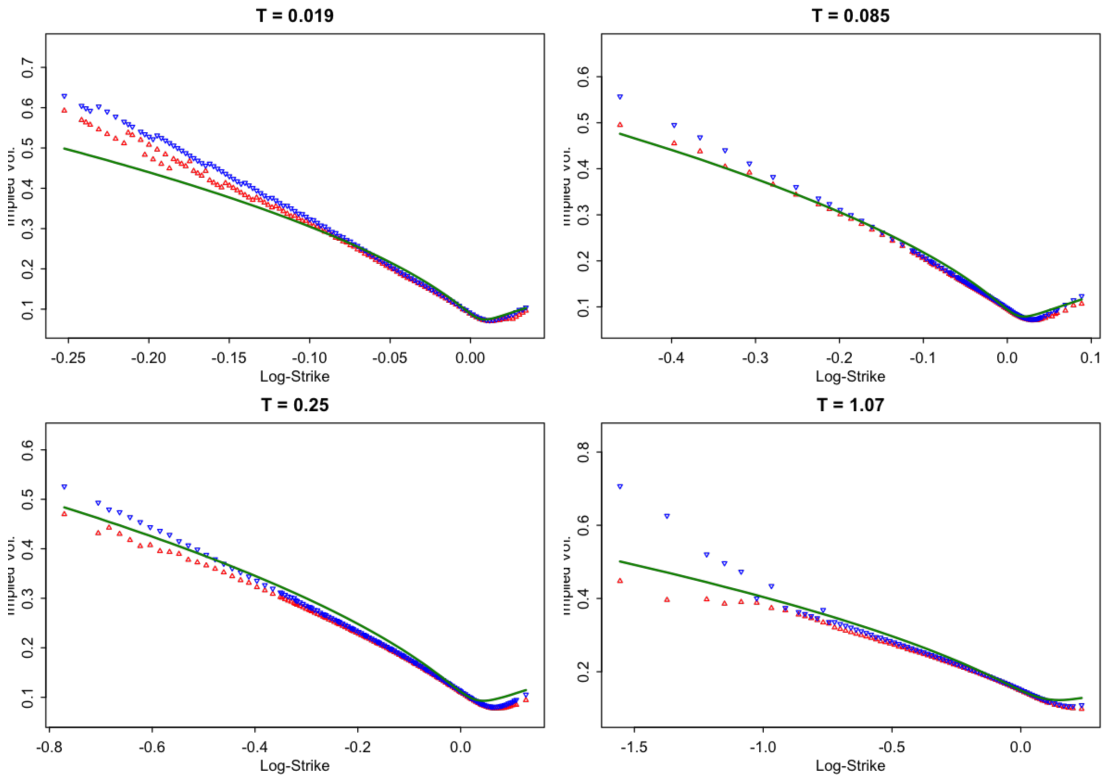

Figure V.2 – Representative SPX volatility smiles as of May 19, 2017, with time to expiration in years. Red and blue points represent bid and ask SPX implied volatilities; green smiles are from the rough Heston model calibrated using the Adams scheme.

We note that on both days for which we plotted smiles, August 14, 2013 and May 19, 2017, the calibrated values of H are small, consistent with the small H estimated from historical data. Thus, it is really true that rough volatility models are amazingly consistent with both historical and implied volatility data.

5 A poor man's rough Heston model

In this section, we present respectively fast and almost instantaneous approximation methods to compute the implied volatility for a given expiry T and parameters (H, ρ, ν). The realized variance of the rough Heston model is given from (2) by

$$\int_0^T V_u du = \int_0^T \xi_0(u) du + \frac{\nu}{\Gamma(H+\frac{3}{2})} \int_0^T (T-u)^{\frac{1}{2}+H} \sqrt{V_u}\, dB_u. \tag{5}$$

The variance of (5) is equal to

$$\nu^2 \int_0^T \frac{(T-s)^{2H+1}}{\Gamma(H+\frac{3}{2})^2} \xi_0(s) ds. \tag{6}$$

This suggests that we approximate the rough Heston smile with the smile generated by a classical Heston-like model (2) with $H = 1/2$ and with a scaled volatility of volatility parameter $\widetilde{\nu}(T)$ matching the variance (6), that is

$$\nu^2 \int_0^T \frac{(T-s)^{2H+1}}{\Gamma(H+\frac{3}{2})^2} \xi_0(s) ds = \widetilde{\nu}(T)^2 \int_0^T (T-s)^2 \xi_0(s) ds.$$

Thus,

$$\widetilde{\nu}(T) = \frac{\nu}{\Gamma(H+\frac{3}{2})} \sqrt{\frac{\int_0^T (T-s)^{2H+1} \xi_0(s) ds}{\int_0^T (T-s)^2 \xi_0(s) ds}}. \tag{7}$$

For each expiry T, the characteristic function formula (3) is then approximated by the classical one

$$\exp\left\{ i\, a\, X_0 + \int_0^T \partial_u h^{(T)}(a, T-u)\, \xi_0(u)\, du \right\} \tag{8}$$

where $h^{(T)}(a,.)$ is now a solution of the classical Riccati equation:

$$\partial_u h^{(T)}(a,u) = -\frac{1}{2}\, a(a+i) + i\,\rho\, \widetilde{\nu}(T)\, a h^{(T)}(a,u) + \frac{1}{2}\, \widetilde{\nu}(T)^2\, (h^{(T)})^2(a,t); \quad h^{(T)}(a,0) = 0.$$

This equation can be solved explicitly as on page 18 of [Gat11]. The solution may be written as

$$h^{(T)}(a,t) = r_-(T) \frac{1 - e^{-A\widetilde{\nu}(T)\, t}}{1 - \frac{r_-(T)}{r_+(T)}\, e^{-A\widetilde{\nu}(T)\, t}}$$

with

$$A = \sqrt{a(a+i) - \rho^2 a^2}; \quad r_\pm(T) = -\frac{1}{\widetilde{\nu}(T)} \left(i\,\rho\, a \pm A \right).$$

On the other hand, for a given expiry T, a poor man's almost instantaneous approximation of the rough Heston characteristic function is obtained by approximating the forward variance curve as flat with $\xi_0(u) = v_0(T)$, $u \geq 0$. In practice, the obvious choice $v_0(T) = \frac{1}{T}\int_0^T \xi_0(s) ds$, the fair value of the variance swap, works fine. In that case, (7) becomes

$$\widetilde{\nu}(T) = \sqrt{\frac{3}{2H+2}}\, \frac{\nu}{\Gamma\left(H+\frac{3}{2}\right)}\, \frac{1}{T^{\frac{1}{2}-H}}. \tag{9}$$

With this choice of forward variance curve, the approximate characteristic function (8) is identical to the characteristic function of the classical Heston model with initial variance $v_0(T)$, mean reversion $\lambda = 0$, correlation ρ and volatility of volatility $\widetilde{\nu}(T)$ given by (9). Option prices under rough Heston may thus be almost instantaneously approximated using any existing implementation of the classical Heston pricing model.

To demonstrate the quality of approximations (7) and (9), in Figure V.3, we have replotted the rough Heston smiles of Figure V.2 generated using the Adams scheme together with those generated using the above approximations. We note that the smiles generated using the approximate characteristic function (8) with approximation (7) appear to be closer to true rough Heston smiles than those generated by the poor man's Heston approximation, especially close to at-the-money. Nevertheless, the poor man's smiles are surprisingly good. An accurate rough Heston approximation with no quants required!

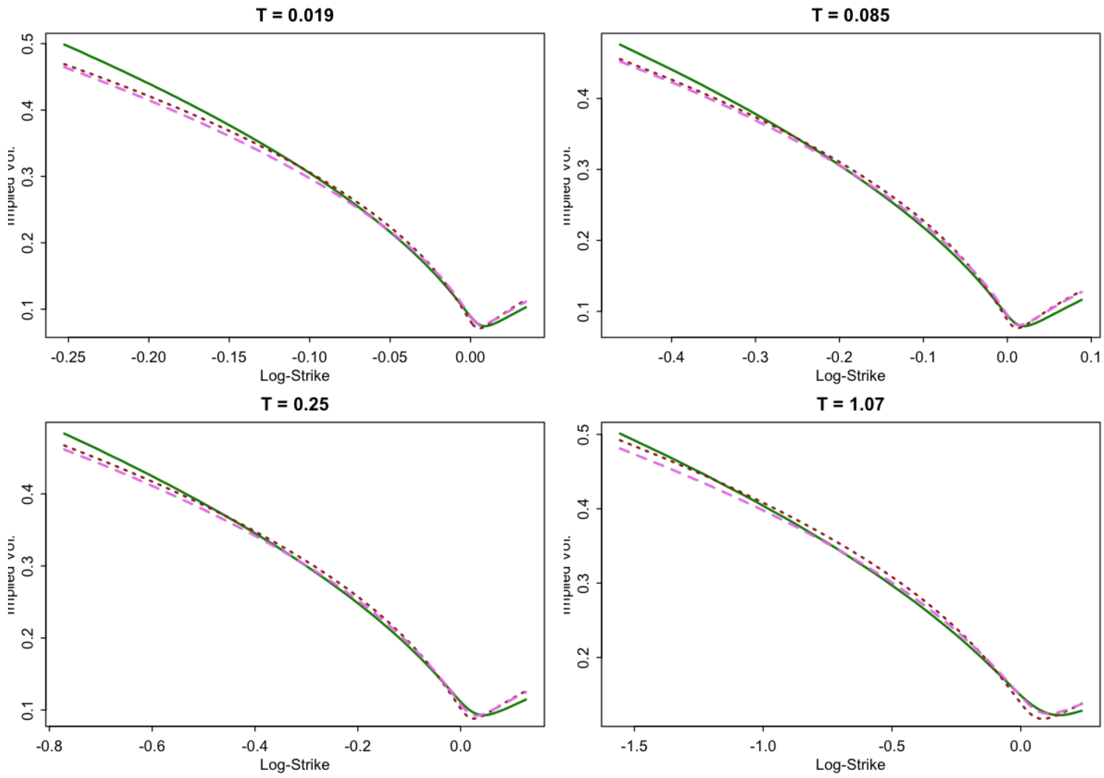

Figure V.3 – Representative SPX volatility smiles as of May 19, 2017, with time to expiration in years. Green smiles are from the rough Heston model calibrated using the Adams scheme; violet dashed smiles are generated using the approximate rough Heston characteristic function (8); the brown dotted smiles are generated from the poor man's existing Heston model with scaled volatility of volatility parameter.

V.A Numerical solution of the fractional Riccati equation

We recall here how to solve fractional ordinary differential equations like (4). This is needed in order to compute the characteristic function. Specifically, again with $\alpha = H + \frac{1}{2}$, let

$$D^{\alpha} h(a,t) = F(a, h(a,t)), \quad h(a,0) = 0. \tag{10}$$

Several schemes for solving (10) numerically are available in the literature. Most of these are based on the idea that (10) implies the following Volterra equation:

$$h(a,t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} F(a, h(a,s)) ds. \tag{11}$$

One way to solve (11) is to use the classical fractional Adams method presented in [DFF04]. The idea goes as follows. Let us write $g(a,t) = F(a,h(a,t))$. Over a regular discrete time-grid with mesh Δ, $0 \le t_0, \ldots \le t_n \le t$, we approximate

$$h(a, t_{k+1}) = \frac{1}{\Gamma(\alpha)} \int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1} g(a,s) ds$$

by

$$\frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \hat{g}(a,s) ds,$$

where

$$\hat{g}(a,t) = \frac{t_{j+1}-t}{t_{j+1}-t_j} \hat{g}(a,t_j) + \frac{t-t_j}{t_{j+1}-t_j} \hat{g}(a,t_{j+1}), \quad t \in [t_j, t_{j+1}], \quad 0 \le j \le k.$$

This corresponds to a trapezoidal discretization and leads to the following scheme:

$$\hat{h}(a, t_{k+1}) = \sum_{0 \le j \le k} a_{j,k+1} F(a, \hat{h}(a,t_j)) + a_{k+1,k+1} F\big(a, \hat{h}(a, t_{k+1})\big), \tag{12}$$

with

$$a_{0,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)} [k^{\alpha+1} - (k-\alpha)(k+1)^\alpha],$$

$$a_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)} [(k-j+2)^{\alpha+1} + (k-j)^{\alpha+1} - 2(k-j+1)^{\alpha+1}]; \quad 1 \le j \le k \tag{13}$$

and

$$a_{k+1,k+1} = \frac{(\Delta t)^\alpha}{\Gamma(\alpha+2)}.$$

However, $\hat{h}(a, t_{k+1})$ being on both sides of (12), this scheme is implicit. Thus, in a first step, we compute a pre-estimation (or *predictor*) of $\hat{h}(a, t_{k+1})$ based on a Riemann sum that we then plug into the trapezoidal quadrature. We define this predictor $\hat{h}^P(a, t_{k+1})$ as follows.

$$\hat{h}^P(a, t_{k+1}) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \tilde{g}(a,s) ds,$$

with

$$\tilde{g}(a,t) = \hat{g}(a,t_j); \quad t \in [t_j, t_{j+1}), \quad 0 \le j \le k.$$

Therefore,

$$\hat{h}^P(a, t_{k+1}) = \sum_{0 \le j \le k} b_{j,k+1} F\big(a, \hat{h}(a,t_j)\big),$$

where

$$b_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+1)} \big((k-j+1)^\alpha - (k-j)^\alpha\big), \quad 0 \le j \le k.$$

Thus, the final explicit numerical scheme is given by

$$\hat{h}(a, t_{k+1}) = \sum_{0 \le j \le k} a_{j,k+1} F(a, \hat{h}(a, t_j)) + a_{k+1,k+1} F\big(a, \hat{h}^P(a, t_j)\big),$$

where the weights $a_{j,k+1}$ are defined in (13).

V.B Call option prices using Fourier techniques

We explain now the way to deal numerically with the Fast Fourier Transform technique of [CM99] for the computation of call option prices in the specific case of the rough Heston model. Recall that the price at time t, $C_t(T, S_0 e^x)$, of the call with expiration time $T > t$ and strike $S_0 e^x$ is related to the characteristic function of the log price X_T through the following expression:

$$C_t(T, S_0 e^x) = \frac{\exp(-\beta x)}{\pi} \int_0^\infty \Re[\psi_t(T, a) e^{-iax}] da, \tag{14}$$

where

$$\psi_t(T, a) = \frac{\phi_t(T, a - (\beta + 1)i)}{\beta^2 + \beta - a^2 + i(2\beta + 1)a}$$

and ϕ_t is defined in (3). Such a method includes the choice of $\beta > 0$ such that

$$\mathbb{E}_t[S_T^{\beta+1}] < \infty. \tag{15}$$

In Chapter III, a sufficient condition for finite moments is given. In particular when $\rho < 0$, the existence of $\beta > 0$ satisfying (15) is guaranteed.

In practice, to apply the Fast Fourier Transform algorithm, we need to tackle the issue of the infinite upper limit of integration in (14) by looking for $a_{max} > 0$ such that

$$\frac{\exp(-\beta x)}{\pi} \int_{a_{max}}^\infty |\psi_t(T, a)| da < \varepsilon,$$

where $\varepsilon > 0$ is the expected truncation error. In Chapter III, it is shown that

$$\Re[\int_t^T D^\alpha h(a - i(\beta + 1), T - u) \xi_t(u) du]$$

is asymptotically dominated as $|a|$ goes to infinity by

$$-|a| \frac{\sqrt{1 - \rho^2}}{\nu \Gamma(1 - \alpha)} \int_t^T (T - u)^{-\alpha} \xi_t(u) du.$$

Hence from (3), it is enough to choose $a_{max} > 0$ such that

$$\frac{\exp(-\beta x)}{\pi} S_t^{\beta+1} \int_{a_{max}}^\infty \frac{\exp\big(-a \frac{\sqrt{1-\rho^2}}{\nu\Gamma(1-\alpha)} \int_t^T (T - u)^{-\alpha} \xi_t(u) du\big)}{|\beta^2 + \beta - a^2 + i(2\beta + 1)a|} da < \varepsilon.$$

CHAPTER VI

Zumbach's effect in rough Heston model

Abstract

Previous literature has identified an effect, dubbed the Zumbach effect, that is nonzero empirically but conjectured to be zero in any conventional stochastic volatility model. Essentially this effect corresponds to the property that past squared returns forecast future volatilities better than past volatilities forecast future squared returns. We provide explicit computations of the Zumbach effect under rough Heston and show that they are consistent with empirical estimates. In agreement with previous conjectures however, the Zumbach effect is found to be negligible in the classical Heston model.

Keywords: Zumbach effect, rough Heston model.

1 Introduction

In a series of papers [BBMZ05, LZ03, ZL01, Zum04, Zum09], Gilles Zumbach and co-authors identified several empirical features of financial time series that are not well replicated by conventional stochastic volatility models. In this paper, we focus on one particular such effect dubbed the *Zumbach effect* in [BDB17].

Denote the true integrated variance from the open to the close of day t by σ_t^2, the open to close return by r_t, and let $\langle \cdot \rangle$ represent a sample average. Then, for $\tau \in \mathbb{R}$, the statistic (6b) of [CB14]

$$\tilde{\mathscr{C}}^{(2)}(\tau) = \langle \left({\sigma_t}^2 - \langle {\sigma_t}^2 \rangle \right) r_{t-\tau}^2 \rangle$$

quantifies (under stationarity assumptions) the covariance of integrated variance with past squared returns. The particular measure of time-reversal asymmetry (TRA) that is found empirically to be positive in [CB14] is given by

$$Z(\tau) := \tilde{\mathscr{C}}^{(2)}(\tau) - \tilde{\mathscr{C}}^{(2)}(-\tau), \quad \tau > 0. \tag{1}$$

In words, the covariance between historical squared returns and future integrated variance is greater than the covariance between historical integrated variance and future squared returns. The following quote from [BDB17] refers to this measure $Z(\tau)$ of TRA:

> Interestingly, all continuous time stochastic volatility models, from the famous CIR-Heston model (Cox et al. 1985, Heston 1993) to the Multifractal Random Walk model alluded to above, obey TRS[1] by construction and therefore *cannot* account for the empirical TRA of financial time series.

In the present paper, we first confirm that $Z(\tau)$ is empirically nonzero. We then compute $Z(\tau)$ explicitly under rough Heston. We show that when the Hurst parameter H of the volatility process is small (H of order 0.1) as established empirically in [GJR18] and confirmed in [BLP16], the Zumbach effect obtained under rough Heston is very consistent with empirical estimates. However, when $H = 1/2$, corresponding to the conventional Heston model, we get that $Z(\tau)$ is indeed numerically absolutely negligible.

Our paper proceeds as follows. In Section 2, we confirm that the Zumbach effect is empirically nonzero. In Section 3, we compute the Zumbach effect under rough Heston. Finally, in Section 4, we show that the rough Heston model is both qualitatively and quantitatively consistent with empirical estimates. Some additional detailed computations are relegated to the appendix.

2 Empirical estimation of the Zumbach effect

For our empirical study, we use opening and closing prices and precomputed realized kernel estimates of intraday (open to close) integrated variance from the Oxford-Man Institute of Quantitative Finance Realized Library from 2000, January 3 to 2018, July 25.[2].

There are 31 indices in the Oxford dataset, as listed in Appendix VI.A. We proceed by computing $\tilde{C}^{(2)}(\tau)$ and $\tilde{C}^{(2)}(-\tau)$ for each of these indices and converting these to correlations by dividing by the relevant sample variances. That is, for each index, we compute

$$\tilde{\rho}(\tau) = \frac{\tilde{C}^{(2)}(\tau)}{\sqrt{\langle(\sigma_t^2 - \langle\sigma_t^2\rangle)^2\rangle \, \langle(r_{t-\tau}^2 - \langle r_{t-\tau}^2\rangle)^2\rangle}}.$$

We then average the $\tilde{\rho}_j$ across the indices j in the dataset to obtain

$$\bar{\rho}(\tau) = \frac{1}{31}\sum_{j=1}^{31} \tilde{\rho}_j(\tau).$$

Finally, corresponding to Equation (25) of [CB14], we further define the integrated difference

$$\Delta(\tau) = \sum_{i=1}^{\tau} \left(\bar{\rho}(i) - \bar{\rho}(-i)\right).$$

In Figure VI.1, we present respectively $\bar{\rho}(\tau)$, $\bar{\rho}(-\tau)$ and $\Delta(\tau)$, reproducing Figure 10 of [CB14], and confirming empirically that the Zumbach effect is nonzero.

[1] Time-reversal symmetry

[2] http://realized.oxford-man.ox.ac.uk/data/download The Oxford-Man Institute of Quantitative Finance Realized Library contains a selection of daily non-parametric estimates of volatility of financial assets, including realized variance and realized kernel estimates. A selection of such estimators is described and their performances compared in, for example, [GO10].

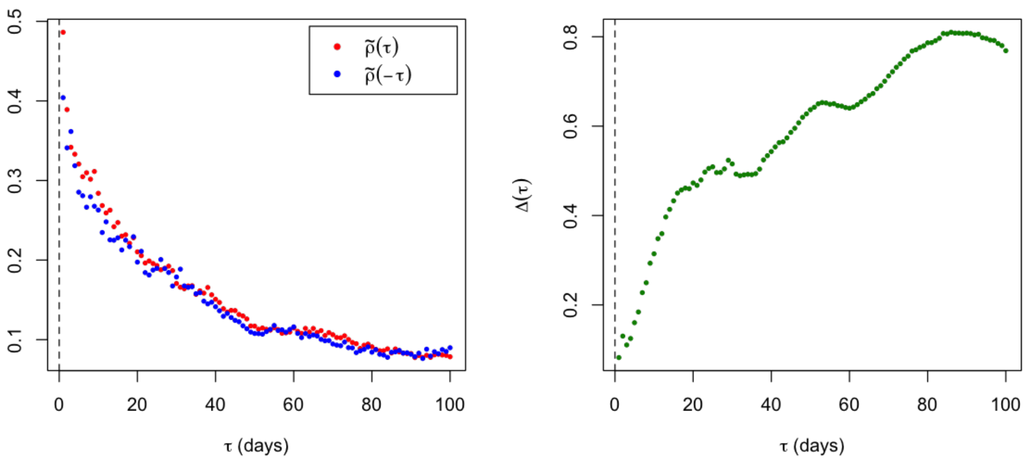

Figure VI.1 – In the left panel, the $\bar{\rho}(\tau)$ are in red, the $\bar{\rho}(-\tau)$ in blue. In the right panel we plot the integrated difference $\Delta(\tau)$.

Remark 1. *Comparing Figure VI.1 with Figure 10 of [CB14], we note that our correlations are in general significantly greater. We attribute this difference to the superior accuracy of the Oxford-Man realized kernel estimates of integrated variance that we use here relative to the Rogers-Satchell estimates computed in [CB14].*

3 The rough Heston model

3.1 Description of the model

We consider the rough Heston model introduced in Chapter II for the price S_t of an asset at time t:

$$dS_t = S_t\sqrt{V_t}dW_t,$$
$$V_t = g_0(t) - \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\lambda V_s ds + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\nu\sqrt{V_s}dB_s. \quad (2)$$

Here $H \in (0,1/2]$ is the Hurst exponent of the volatility, $\lambda > 0$ is the mean reversion parameter, $\nu > 0$ is the volatility of volatility parameter and (W,B) is a ρ-correlated Brownian motion with $\rho \in [-1,1]$. The function g_0 is assumed to be continuous and is linked to the forward variance curve $\xi_0(t) = \mathbb{E}[V_t]$ as follows:

$$g_0(t) = \xi_0(t) + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\lambda\xi_0(s)ds. \quad (3)$$

Note that in Chapter IV, a general condition on g_0 is given to guarantee weak existence and uniqueness for the solution of the equation defining the rough Heston model.

In Chapters II and III, it is shown that there exists a semi-closed form expression for the characteristic function, just as in the classical Heston case, and that explicit hedging strategies can be derived. Fast and accurate option pricing is also possible, see [GR18]. Furthermore, the rough Heston model displays the rough behavior of the volatility observed empirically in [GJR18]. More precisely, the variance process V admits Hölder continuous paths with regularity strictly less than H. In addition to the fit to historical data, it is shown in Chapter V that with suitably calibrated parameters (H, ν, ρ and λ), the rough Heston model typically fits the SPX volatility surface remarkably well.

3.2 The Zumbach effect under rough Heston: explicit computation

We provide in this section an explicit formula for the Zumbach effect in the rough Heston model (Theorem 1). We start with a discussion about the use of correlations or covariances when computing the Zumbach effect under rough Heston.

3.2.1 Correlations versus covariances

From a theoretical viewpoint, approximating theoretical quantities such as covariances and correlations by sample values makes sense only provided the underlying dynamics can be considered stationary. In the context of the rough Heston model (2), this would imply that the parameter λ should be large enough with respect to the observation time scale. However, whether rough volatility models are estimated under $\mathbb{P}$ or $\mathbb{Q}$, λ is typically found to be small relative to this observation time scale, see Chapter V. In this case, under rough Heston, the very notion of the Zumbach effect may appear somehow ill-defined.

It turns out, as will be seen in Section 3.2.2, that under rough Heston, the Zumbach effect $Z(\tau)$ expressed as a difference of covariances (1), does not depend asymptotically on λ. This is in contrast to the effect $Z^{\text{Correl}}(\cdot)$ expressed in terms of corrrelations as in Proposition 1. Consequently, we choose to focus on covariances and express the Zumbach effect in terms of the covariances $\tilde{C}^{(2)}$ rather than the correlations $\tilde{\rho}$. For the sake of completeness, computations based on correlation in the stationary regime are presented in Appendix VI.C.

3.2.2 Computation of the Zumbach effect

Denote the length of the trading day by δ. Under the rough Heston model, the open to close (log-)return is given by[3]

$$r_t = \int_{t-\delta}^{t} \sqrt{V_s} dW_s,$$

and the daily integrated variance by

$$\sigma_t^2 = \int_{t-\delta}^{t} V_s ds.$$

[3]For simplicity, we take only the martingale part of the log-returns.

Our aim is to prove that in the rough Heston model, the counterpart of $Z(\tau) = \tilde{\mathscr{C}}^{(2)}(\tau) - \tilde{\mathscr{C}}^{(2)}(-\tau)$ is positive for $\tau = k\delta$ with $k \in \mathbb{N}_{>0}$. Hence, we write

$$Z_t(k) = \mathrm{Cov}[r_t^2, \sigma_{t+k\delta}^2] - \mathrm{Cov}[r_{t+k\delta}^2, \sigma_t^2], \quad k \in \mathbb{N}_{>0}, \quad t \geq \delta.$$

By Itô's isometry, $\mathbb{E}\left[r_t^2\right] = \mathbb{E}\left[\sigma_t^2\right]$ for any time t so

$$Z_t(k) = \mathbb{E}[r_t^2\, \sigma_{t+k\delta}^2] - \mathbb{E}[r_{t+k\delta}^2\, \sigma_t^2].$$

Applying Itô's formula, we get

$$r_t^2 = 2\int_{t-\delta}^{t} \sqrt{V_s}\left(\int_{t-\delta}^{s} \sqrt{V_u} dW_u\right) dW_s + \sigma_t^2.$$

This together with the fact that $\mathbb{E}\left[r_{t+k\delta}^2\, \sigma_t^2\right] = \mathbb{E}\left[\mathbb{E}\left[\, r_{t+k\delta}^2 \middle| \mathscr{F}_t\right] \sigma_t^2\right] = \mathbb{E}[\sigma_{t+k\delta}^2\, \sigma_t^2]$ leads to

$$Z_t(k) = 2\,\mathbb{E}\left[\sigma_{t+k\delta}^2 \int_{t-\delta}^{t} \sqrt{V_s}\left(\int_{t-\delta}^{s} \sqrt{V_u} dW_u\right) dW_s\right]. \tag{4}$$

Substituting (3) into (2) with $\alpha = H + \frac{1}{2}$ gives

$$V_t = \xi_0(t) - \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda\, (V_s - \xi_0(s))\, ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \nu \sqrt{V_s} dB_s.$$

From Lemma 1, the solution is given by

$$V_t = \xi_0(t) + \int_0^t f^{\alpha,\lambda}(t-s) \frac{\nu}{\lambda} \sqrt{V_s} dB_s \tag{5}$$

where $f^{\alpha,\lambda}(x) = \lambda x^{\alpha-1} \sum_{k \geq 0} \frac{(-\lambda x^\alpha)^k}{\Gamma(\alpha(k+1))}$ is the Mittag-Leffler density function. It follows that the future integrated variance satisfies

$$\begin{aligned}
\sigma_{t+k\delta}^2 &= \int_{t+(k-1)\delta}^{t+k\delta} V_s\, ds \\
&= \int_{t+(k-1)\delta}^{t+k\delta} \xi_0(s)\, ds + \int_0^{t+k\delta} F^{\alpha,\lambda}(t+k\delta-s) \frac{\nu}{\lambda} \sqrt{V_s} dB_s \\
&- \int_0^{t+(k-1)\delta} F^{\alpha,\lambda}(t+(k-1)\delta-s) \frac{\nu}{\lambda} \sqrt{V_s} dB_s,
\end{aligned}$$

where $F^{\alpha,\lambda}(x) = \int_0^x f^{\alpha,\lambda}(s) ds$. Consequently (4) may be rewritten as

$$Z_t(k) = 2\frac{\rho\nu}{\lambda} \int_{t-\delta}^{t} \left(F^{\alpha,\lambda}(t+k\delta-s) - F^{\alpha,\lambda}(t+(k-1)\delta-s)\right) \mathbb{E}\left[V_s \int_{t-\delta}^{s} \sqrt{V_u} dW_u\right] ds.$$

Again using (5), we have that for $s \geq t-\delta$,

$$\mathbb{E}\left[V_s \int_{t-\delta}^{s} \sqrt{V_u} dW_u\right] = \frac{\rho\nu}{\lambda} \int_{t-\delta}^{s} f^{\alpha,\lambda}(s-u)\, \xi_0(u)\, du.$$

Thus,

$$Z_t(k) = 2\,\frac{(\rho\,\nu)^2}{\lambda^2}\int_0^{\delta}\left(F^{\alpha,\lambda}(s+k\delta)-F^{\alpha,\lambda}(s+(k-1)\delta)\right)\int_0^{\delta-s} f^{\alpha,\lambda}(u)\xi_0(t-s-u)\,du\,ds, \tag{6}$$

which is positive if ρ is different from zero. Using the fact that

$$F^{\alpha,\lambda}(x)\underset{x\to 0}{\sim}\frac{\lambda\,x^{\alpha}}{\Gamma(\alpha+1)},$$

together with the dominated convergence theorem, we derive the following result.

Theorem 1. *Assume that ρ is nonzero and that the forward variance curve is continuous. Then $Z_t(k)>0$ and as δ goes to zero,*

$$Z_t(k)\underset{\delta\to 0}{\sim} 2(\rho\,\nu)^2\,\delta^{2\alpha+1}\,g_\alpha(k)\,\xi_0(t),\quad k\in\mathbb{N}_{>0},\quad t>0,$$

with $g_\alpha(k)=\frac{1}{\Gamma(\alpha+1)^2}\int_0^1((k+s)^{\alpha}-(k+s-1)^{\alpha})\,(1-s)^{\alpha}ds$.

We see that this measure of the Zumbach effect is indeed independent of λ. It is also independent of t in the flat forward variance curve case.

4 Numerical results

To compare model computations with empirical estimates, we adopt the following model parameters typical of calibrations to the SPX implied volatility surface:

$$\rho=-0.7,\quad \nu=0.45,\quad H=0.05,\quad \lambda=0.3.$$

We assume a flat forward variance curve setting $\xi_0(t)=0.025$, the approximate sample mean of ${\sigma_t}^2$.

In Figure VI.2, we superimpose empirical estimates $Z(\tau)=Z(k\delta)$ and model computations $Z_t(k)$ for SPX (which do not depend on t here). Although model computations are somewhat higher than empirical estimates, we argue that this nevertheless represents good agreement between model and data. One factor no doubt contributing to the discrepancy is that we expect volatility of volatility and correlation under $\mathbb{Q}$ to be more extreme than their equivalents under $\mathbb{P}$.

4.1 Dependence on H

We now examine the dependence of $Z_t(k)$ on the Hurst exponent H. We already showed that under rough Heston with reasonable parameters, the Zumbach effect is consistent with empirical estimates. In contrast, when $H=1/2$, we see from Figure VI.3 that the Zumbach effect is negligible. Indeed, from Theorem 1, $Z_t(k)$ is of order $\delta^{2\alpha+1}=\delta^{2H}$ for small δ. When $H=1/2$, $Z_t(k)\sim\delta$ becomes very small, whereas as $H\to 0$, that is when volatility is rough, the Zumbach effect remains significant.

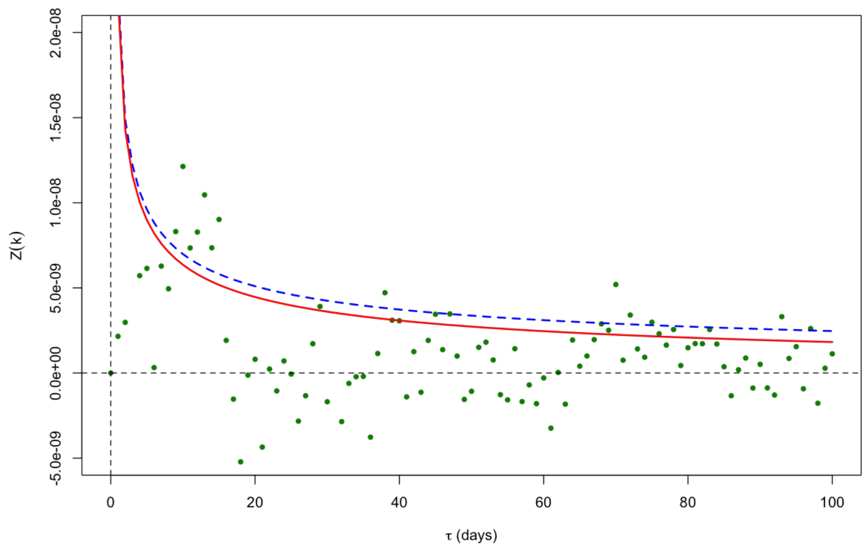

Figure VI.2 – With $\tau = k\delta$, the green points are the empirical estimates of $Z(\tau)$, the solid red line is the model computation (6) of $Z_t(k)$, the dashed blue line is the small δ approximation from Theorem 1 to $Z_t(k)$.

VI.A List of indices in the Oxford-Man Institute of Quantitative Finance Realized Library

The following table lists all of the index tickers included in the Oxford-Man Institute of Quantitative Finance Realized Library together with index descriptions.

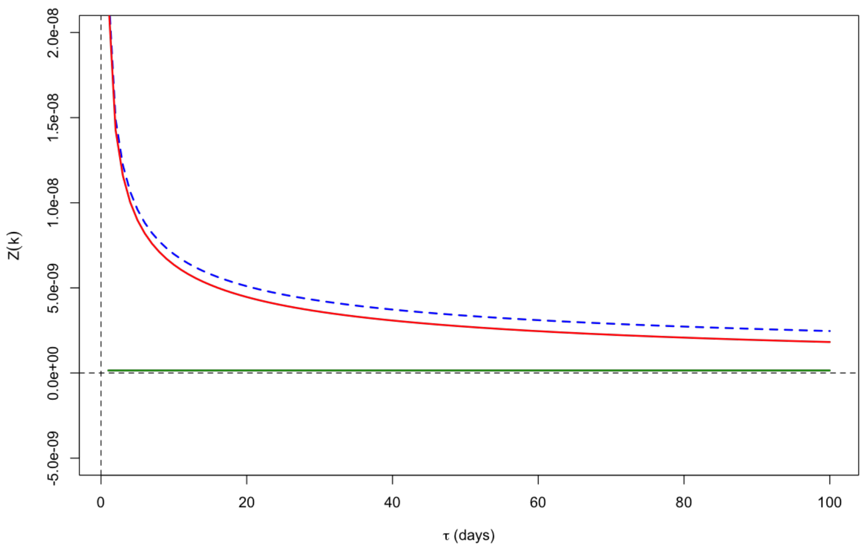

Figure VI.3 – The solid red line is $Z_t(k)$ computed with $H = 0.05$, the blue dashed line is the approximation from Theorem 1, and the green line close to the x-axis is $Z_t(k)$ with $H = 1/2$. We see that the effect is negligible when $H = 1/2$.

Index ticker	Index description
.AEX	Amsterdam Exchange Index
.AORD	All Ordinaries Index
.BFX	BEL 20 Index
.BSESN	S&P Bombay Stock Exchange SENSEX Index
.BVLG	Euronext PSI General Index
.BVSP	BOVESPA Index
.DJI	Dow Jones Industrial Average
.FCHI	CAC 40
.FTMIB	FTSE MIB Index
.FTSE	FTSE 100 Index
.GDAXI	DAX Index
.GSPTSE	S&P/TSX Composite Index
.HSI	Hang Seng Index
.IBEX	IBEX 35 Index
.IXIC	Nasdaq Composite Index
.KS11	KOSDAQ Composite Index
.KSE	Karachi Stock Exchange 100 Index
.MXX	Mexican Bolsa IPC Index
.N225	Nikkei 225 Index
.NSEI	NIFTY 50 Index
.OMXC20	OMX Copenhagen 20 Index
.OMXHPI	OMX Helsinki All-Share Index
.OMXSPI	OMX Stockholm All-Share Index
.OSEAX	Oslo Børs All-Share Index
.RUT	Russell 2000 Index
.SMSI	Madrid General Index
.SPX	S&P 500 Index
.SSEC	Shanghai Composite Index
.SSMI	Swiss Market Index
.STI	Straits Times Index
.STOXX50E	Euro STOXX 50 Index

VI.B Proof of (5)

The following technical lemma slightly extends Proposition 4.10 in Chapter I.

Lemma 1. *The process V is solution of the following rough stochastic differential equation*

$$V_t = \xi_0(t) - \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda \left(V_s - \xi_0(s)\right) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \nu \sqrt{V_s} dB_s$$

if and only if it is solution of

$$V_t = \xi_0(t) + \int_0^t f^{\alpha,\lambda}(t-s) \frac{\nu}{\lambda} \sqrt{V_s} dB_s.$$

Proof. Suppose

$$V_t = \xi_0(t) + \int_0^t f^{\alpha,\lambda}(t-s) \frac{\nu}{\lambda} \sqrt{V_s} dB_s.$$

Then[4] using fractional integration of order $1-\alpha$ (denoted by $I^{1-\alpha}$), the properties of the Mittag-Leffler density and the stochastic Fubini theorem, this is equivalent to

$$\begin{aligned}
I^{1-\alpha} V_t &= I^{1-\alpha}\xi_0(t) + \frac{\nu}{\lambda} \int_0^t I^{1-\alpha} f^{\alpha,\lambda}(t-s) \sqrt{V_s} dB_s \\
&= I^{1-\alpha}\xi_0(t) + \frac{\nu}{\lambda} \int_0^t \lambda \left(1 - F^{\alpha,\lambda}(t-s)\right) \sqrt{V_s} dB_s \\
&= I^{1-\alpha}\xi_0(t) + \nu \int_0^t \sqrt{V_s} dB_s - \nu \int_0^t dB_s \int_s^t f^{\alpha,\lambda}(u-s) \sqrt{V_s} du \\
&= I^{1-\alpha}\xi_0(t) + \nu \int_0^t \sqrt{V_s} dB_s - \nu \int_0^t du \int_0^u f^{\alpha,\lambda}(u-s) \sqrt{V_s} dB_s \\
&= I^{1-\alpha}\xi_0(t) + \nu \int_0^t \sqrt{V_s} dB_s - \lambda \int_0^t \left(V_u - \xi_0(u)\right) du.
\end{aligned}$$

Finally, applying fractional differentiation of order $1-\alpha$ together with the stochastic Fubini theorem we deduce the result. □

VI.C The Zumbach effect in terms of correlations in the stationary regime

We now discuss the Zumbach effect in terms of correlations in the stationary regime, that is when t goes to infinity. In particular, we suppose that $\xi_0(t)$ satisfies

$$\xi_0(t) \longrightarrow \xi_0(\infty), \tag{7}$$

[4] For definitions and properties of fractional integration and differentiation and of the Mittag-Leffler density $f^{\alpha,\lambda}$, see for example Appendices A3 and A4 of Chapter I

as t goes infinity for some $\xi_0(\infty) > 0$. From Theorem 1, we have that for small δ, $Z_t(k)$ is equivalent to

$$2(\rho\nu)^2\delta^{2\alpha+1}g_\alpha(k)\xi_0(\infty).$$

Moreover, from Appendix VI.D, the limit of $\mathbb{E}[r_t^4]$ as t goes to infinity is

$$\xi_0(\infty)\frac{12\rho^2\nu^2}{\lambda^2}\int_0^\delta F^{\alpha,\lambda}(s)F^{\alpha,\lambda}(\delta-u)du + 3\xi_0(\infty)^2\delta^2 + \frac{3\nu^2}{\lambda^2}\xi_0(\infty)\int_0^\delta F^{\alpha,\lambda}(u)^2du$$
$$+\frac{6\nu^2}{\lambda^2}\xi_0(\infty)\int_0^\infty\left(\int_0^\delta\left(F^{\alpha,\lambda}(s+u)-F^{\alpha,\lambda}(u)\right)f^{\alpha,\lambda}(s+u)ds\right)du.$$

This limit is equivalent for small δ to

$$3\xi_0(\infty)^2\delta^2 + \frac{3\nu^2}{\lambda^2}\xi_0(\infty)\delta^2\int_0^\infty f^{\alpha,\lambda}(s)^2ds.$$

In the same way, from Appendix VI.D, the limit of $\mathrm{Var}[\sigma_t^2]$ as t goes to infinity is

$$\frac{\nu^2}{\lambda^2}\xi_0(\infty)\int_0^\infty\left(F^{\alpha,\lambda}(s+\delta)-F^{\alpha,\lambda}(s)\right)^2ds + \frac{\nu^2}{\lambda^2}\xi_0(\infty)\int_0^\delta F^{\alpha,\lambda}(s)^2ds,$$

which is equivalent to

$$\frac{\nu^2}{\lambda^2}\xi_0(\infty)\delta^2\int_0^\infty f^{\alpha,\lambda}(s)^2ds.$$

Let us now define the correlation based Zumbach effect $Z^{\mathrm{Correl}}(k)$ by

$$Z^{\mathrm{Correl}}(k) = \frac{\lim\limits_{t\to\infty} Z_t(k)}{\sqrt{\lim\limits_{t\to\infty}\mathrm{Var}[\sigma_t^2]\mathrm{Var}[r_t]}}.$$

From previous computations, we deduce the following proposition.

Proposition 1. *We have*

$$Z^{\mathrm{Correl}}(k) \underset{\delta\to 0}{\sim} \frac{2(\rho\nu)^2\sqrt{\xi_0(\infty)}}{\sqrt{\frac{\nu^2}{\lambda^2}\int_0^\infty f^{\alpha,\lambda}(s)^2ds}\sqrt{2\xi_0(\infty)+\frac{3\nu^2}{\lambda^2}\int_0^\infty f^{\alpha,\lambda}(s)^2ds}}\delta^{2\alpha-1}g_\alpha(k).$$

VI.D Variance computations

We compute in this section $\mathrm{Var}[\sigma_t^2]$ and $\mathrm{Var}[r_t^2]$. Using (5) together with the stochastic Fubini theorem, we have that

$$\sigma_t^2 = \int_{t-\delta}^t \xi_0(s)ds + \int_0^t \frac{\nu}{\lambda}F^{\alpha,\lambda}(t-s)\sqrt{V_s}dB_s - \int_0^{t-\delta}\frac{\nu}{\lambda}F^{\alpha,\lambda}(t-\delta-s)\sqrt{V_s}dB_s. \quad (8)$$

Hence,

$$\mathrm{Var}[\sigma_t^2] = \frac{\nu^2}{\lambda^2}\int_0^{t-\delta}\left(F^{\alpha,\lambda}(s+\delta)-F^{\alpha,\lambda}(s)\right)^2\xi_0(t-\delta-s)ds + \frac{\nu^2}{\lambda^2}\int_0^\delta F^{\alpha,\lambda}(s)^2\xi_0(t-s)ds.$$

In order to compute $\mathrm{Var}[r_t^2]$, we need to get $\mathbb{E}[r_t^4]$. Note that by Itô's formula,

$$\mathbb{E}[r_t^4] = 6\int_{t-\delta}^{t} \mathbb{E}\left[\left(\int_{t-\delta}^{s} \sqrt{V_u}dW_u\right)^2 V_s\right] ds.$$

Using again Itô's formula, we get that $\mathbb{E}\left[\left(\int_{t-\delta}^{s} \sqrt{V_u}dW_u\right)^2 V_s\right]$ is equal to

$$2\mathbb{E}\left[V_s\int_{t-\delta}^{s} \sqrt{V_u}\int_{t-\delta}^{u} \sqrt{V_w}dW_w dW_u\right] + \mathbb{E}\left[V_s\int_{t-\delta}^{s} V_u du\right]. \tag{9}$$

From (5), the first term in (9) is given by

$$\frac{2\rho\nu}{\lambda}\int_{t-\delta}^{s} f^{\alpha,\lambda}(s-u)\mathbb{E}\left[V_u\int_{t-\delta}^{u} \sqrt{V_w}dW_w\right] du,$$

which is equal to

$$\frac{2\rho^2\nu^2}{\lambda^2}\int_{t-\delta}^{s} f^{\alpha,\lambda}(s-u)\left(\int_{t-\delta}^{u} f^{\alpha,\lambda}(u-w)\xi_0(w)dw\right) du.$$

Hence the first term in (9) is equal to

$$\frac{2\rho^2\nu^2}{\lambda^2}\int_{0}^{s-t+\delta} f^{\alpha,\lambda}(u)\left(\int_{0}^{s-u-t+\delta} f^{\alpha,\lambda}(w)\xi_0(s-u-w)dw\right) du.$$

Moreover, similarly to (8),

$$\int_{t-\delta}^{s} V_u du = \int_{t-\delta}^{s} \xi_0(u)du + \int_0^s \frac{\nu}{\lambda}F^{\alpha,\lambda}(s-u)\sqrt{V_u}dB_u - \int_0^{t-\delta} \frac{\nu}{\lambda}F^{\alpha,\lambda}(t-\delta-u)\sqrt{V_u}dB_u.$$

Therefore the second term of (9) is given by

$$\xi_0(s)\int_0^{s-t+\delta} \xi_0(u+t-\delta)du + \frac{\nu^2}{\lambda^2}\int_0^{s-t+\delta} f^{\alpha,\lambda}(u)F^{\alpha,\lambda}(u)\xi_0(s-u)du$$
$$+\frac{\nu^2}{\lambda^2}\int_0^{t-\delta}\left(F^{\alpha,\lambda}(s-t+\delta+u) - F^{\alpha,\lambda}(u)\right)f^{\alpha,\lambda}(s-t+\delta+u)\xi_0(t-\delta-u)du.$$

Consequently, $\mathbb{E}[r_t^4]$ is equal to

$$\frac{12\rho^2\nu^2}{\lambda^2}\int_0^{\delta}\int_0^{s} f^{\alpha,\lambda}(u)\left(\int_0^{s-u} f^{\alpha,\lambda}(w)\xi_0(s-u+t-\delta-w)dw\right)duds$$
$$+6\int_0^{\delta}\xi_0(s+t-\delta)\int_0^s \xi_0(u+t-\delta)duds + \frac{6\nu^2}{\lambda^2}\int_0^{\delta}\int_0^{s} f^{\alpha,\lambda}(u)F^{\alpha,\lambda}(u)\xi_0(s+t-\delta-u)duds$$
$$+\frac{6\nu^2}{\lambda^2}\int_0^{t-\delta}\left(\int_0^{\delta}\left(F^{\alpha,\lambda}(s+u) - F^{\alpha,\lambda}(u)\right)f^{\alpha,\lambda}(s+u)ds\right)\xi_0(t-\delta-u)du.$$

Part V

Short-term behavior of the at-the-money implied volatility under rough volatility models

CHAPTER VII

Short-term at-the-money asymptotics under stochastic volatility models

Abstract

A small-time Edgeworth expansion of the density of an asset price is given under a general stochastic volatility model, from which asymptotic expansions of put option prices and at-the-money implied volatilities follow. A limit theorem for at-the-money implied volatility skew and curvature is also given as a corollary. The rough Bergomi model is treated as an example.

1 Introduction

A stochastic volatility model is an extension of the Black-Scholes model that incorporates an empirical evidence that the volatility of an asset price is not constant in its time-series data as well as in its option price data. The Heston and SABR models among others are popular in financial practices owing to (semi-)analytic (approximation) formulas for the vanilla option prices or the option-implied volatilities. See e.g., [Gat11] for a practical guide on stochastic volatility modeling.

Recently, attracting much attention is a class of stochastic volatility models where the volatility is driven by a fractional Brownian motion. This is due to their consistency to a power law of the term structure in the implied volatility skew which has been empirically recognized; see [ALV07, BFG16, FZ17, Fuk11, Fuk17, GS17, GJR18, GJR14]. To be consistent, the fractional Brownian motion must be correlated with a Brownian motion driving the asset price and its Hurst parameter must be smaller than 1/2. The latter means in particular that the volatility path is rougher than a Brownian motion and so, this class of the models is often referred as the rough volatility models. Since the models do not admit of explicit expressions for option prices or implied volatilities, the above mentioned consistency has been discussed through asymptotic analyses.

The aim of this paper is to provide a general framework under which the short-term asymptotics of the at-the-money implied volatility is studied. The framework is for a general continuous stochastic volatility model. The rough Bergomi model introduced by [BFG16] is treated as an example. The asymptotic expansion of the at-the-money implied volatility is given up to the second-order, while the first order expansion was already given in [Fuk17] by a different method. For the SABR model Osajima [Osa07] gave the expansion based on the Watanabe-Yoshida theory; see e.g., [KT⁺03, Yos92]. The same expansion formula was also obtained in Medvedev and Scaillet [MS07] by a formal computation. Friz et al. [FGP17] derived the asymptotic skew and curvature of the implied volatility that correspond to the first and the second order terms by assuming the asymptotic behavior of the density function of the underlying asset price. Here, we introduce a novel approach based on a conditional Gaussianity of the stochastic volatility model to prove the validity of a second order density expansion for a general stochastic volatility model. From this density expansion follow expansions of the option prices and the implied volatility as well as the asymptotic skew and curvature formula. In contrast to [KT⁺03, Osa07, Yos92], we do not rely on the Malliavin calculus, which enables us to treat effectively the rough volatility models. In contrast to the elementary method of [Fuk17], our approach can be extended to higher-order expansions without any additional theoretical difficulty. We choose the square root of the forward variance, that is, the fair strike of a variance swap, as the leading term of our asymptotic expansion, while a recent work [AS] studies the difference between the implied volatility and the fair strike of a volatility swap in terms of the Malliavin derivatives.

The paper is organized as follows. In Section 2, we describe the model framework. In Section 3, we derive the asymptotic expansions of the characteristic function, the density function, the put option prices and the at-the-money implied volatility function. In Section 4, we derive the asymptotic behavior of the at-the-money implied volatility skew and curvature. In Section 5, we show that the rough Bergomi model fits the framework and compute the coefficients of the expansion for this particular model.

2 Framework

2.1 Assumptions

Let $(\Omega, \mathscr{F}, P)$ be a probability space equipped with a filtration $\{\mathscr{F}_t; t \in \mathbb{R}\}$ satisfying the usual assumptions. A log price process Z is assumed to follow

$$\mathrm{d}Z_t = r\mathrm{d}t - \frac{1}{2}v_t\mathrm{d}t + \sqrt{v_t}\mathrm{d}B_t$$

under an equivalent measure Q, where $r \in \mathbb{R}$ stands for an interest rate and v is a positive continuous process adapted to a smaller filtration $\{\mathscr{G}_t; t \in \mathbb{R}\}$, of which the square root is called the volatility of Z. The Brownian motion B is decomposed as

$$\mathrm{d}B_t = \rho_t\mathrm{d}W_t + \sqrt{1-\rho_t^2}\mathrm{d}W_t',$$

where W' is an $\{\mathscr{F}_t\}$-Brownian motion independent of $\mathscr{G}_t$ for all $t \in \mathbb{R}$, W is a $\{\mathscr{G}_t\}$-Brownian motion and ρ is a progressively measurable processes with respect to $\{\mathscr{G}_t\}$ and taking values in $[-1,1]$. A typical situation for stochastic volatility models, including the Heston, SABR and rough Bergomi models, is that (W, W') is a two dimensional $\{\mathscr{F}_t\}$-Brownian motion and $\{\mathscr{G}_t\}$ is the filtration generated by W, that is,

$$\mathscr{G}_t = \mathscr{N} \vee \sigma(W_s - W_r; r \le s \le t),$$

where $\mathscr{N}$ is the null sets of $\mathscr{F}$.

An arbitrage-free price $p(K,\theta)$ of a put option at time 0 with strike $K > 0$ and maturity $\theta > 0$ is given by

$$p(K,\theta) = e^{-r\theta} E^Q[(K - \exp(Z_\theta))_+ | \mathscr{F}_0] = e^{-r\theta} \int_0^K Q(\log x \ge Z_\theta | \mathscr{F}_0) \mathrm{d}x.$$

Denote by E_0 and $\|\cdot\|_p$ respectively the expectation and the L^p norm under the regular conditional probability measure of Q given $\mathscr{F}_0$, of which the existence is assumed. We impose the following technical condition: for any $p > 0$,

$$\sup_{\theta \in (0,1)} \left\| \frac{1}{\theta} \int_0^\theta v_t \mathrm{d}t \right\|_p < \infty, \quad \sup_{\theta \in (0,1)} \left\| \left\{ \frac{1}{\theta} \int_0^\theta v_t (1 - \rho_t^2) \mathrm{d}t \right\}^{-1} \right\|_p < \infty. \tag{1}$$

The forward variance curve $v^0(t)$ is defined by

$$v^0(t) = E_0[v_t] = E^Q[v_t | \mathscr{F}_0].$$

Let

$$\sigma_0(\theta) = \sqrt{\int_0^\theta v^0(t) \mathrm{d}t}.$$

Changing variable as

$$x = F \exp(\zeta \sigma_0(\theta)), \quad F = \exp(r\theta + Z_0),$$

we have

$$\frac{p(Fe^{z\sigma_0(\theta)}, \theta)}{F\sigma_0(\theta)} = e^{-r\theta} \int_{-\infty}^z Q(\zeta \ge X_\theta | \mathscr{F}_0)\, e^{\sigma_0(\theta)\zeta} \mathrm{d}\zeta,$$

where

$$X_\theta = -\frac{1}{2\sigma_0(\theta)} \langle M \rangle_\theta + \frac{1}{\sigma_0(\theta)} M_\theta, \quad M_\theta = \int_0^\theta \sqrt{v_t} \mathrm{d}B_t, \quad \langle M \rangle_\theta = \int_0^\theta v_t \mathrm{d}t.$$

We assume the following asymptotic structure: there exists a family of random vectors

$$\left\{ (M_\theta^{(0)}, M_\theta^{(1)}, M_\theta^{(2)}, M_\theta^{(3)}); \theta \in (0,1) \right\}$$

such that

1. the law of $M_\theta^{(0)}$ is standard normal for all $\theta > 0$,

2.
$$\sup_{\theta\in(0,1)} \|M_\theta^{(i)}\|_p < \infty, \quad i = 1,2,3 \tag{2}$$
for all $p > 0$ and

3. for some $H \in (0,1/2]$ and $\epsilon \in (0,H)$,
$$\begin{aligned}
&\lim_{\theta\to 0} \theta^{-2H-2\epsilon} \left\| \frac{M_\theta}{\sigma_0(\theta)} - M_\theta^{(0)} - \theta^H M_\theta^{(1)} - \theta^{2H} M_\theta^{(2)} \right\|_{1+\epsilon} = 0, \\
&\lim_{\theta\to 0} \theta^{-H-2\epsilon} \left\| \frac{\langle M\rangle_\theta}{\sigma_0(\theta)^2} - 1 - \theta^H M_\theta^{(3)} \right\|_{1+\epsilon} = 0.
\end{aligned} \tag{3}$$

Further, we assume the existence of the derivatives
$$\begin{aligned}
a_\theta^{(i)}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left\{E_0[M_\theta^{(i)}|M_\theta^{(0)} = x]\phi(x)\right\}, \quad i = 1,2,3, \\
b_\theta(x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left\{E_0[M_\theta^{(1)}|M_\theta^{(0)} = x]\phi(x)\right\} \\
c_\theta(x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left\{E_0[|M_\theta^{(1)}|^2|M_\theta^{(0)} = x]\phi(x)\right\}
\end{aligned} \tag{4}$$
in the Schwartz space (i.e., the space of the rapidly decreasing smooth functions), where ϕ is the standard normal density.

2.2 Regular stochastic volatility models

Here we briefly discuss that regular stochastic volatility models satisfy all the above assumptions. Let us consider the volatility process $v_t = v(X_t)$, where X is a Markov process satisfying a stochastic differential equation
$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + c(X_t)\mathrm{d}W_t$$
and v is a smooth positive function defined on the state space of X. Let $\rho \in (-1,1)$ be a constant and $\{\mathscr{G}_t\}$ be the augmented filtration generated by W. We assume (1), which is satisfied in the usual cases including the log-normal SABR and Heston models. Denote by L the generator of X. Put $f = \sqrt{v}$, $g = f'c$ and $h = v'c$. Then, by Itô's formula, we have
$$\begin{aligned}
M_\theta &= f(X_0)B_\theta + \int_0^\theta \int_0^t g(X_s)\mathrm{d}W_s\mathrm{d}B_t + \int_0^\theta \int_0^t Lf(X_s)\mathrm{d}s\mathrm{d}B_t, \\
\langle M\rangle_\theta &= v(X_0)\theta + \int_0^\theta \int_0^t h(X_s)\mathrm{d}W_s\mathrm{d}t + \int_0^\theta \int_0^t Lv(X_s)\mathrm{d}s\mathrm{d}t.
\end{aligned}$$
Let $\bar{B}_t^\theta = \theta^{-1/2}B_{\theta t}$, $\bar{W}_t^\theta = \theta^{-1/2}W_{\theta t}$ and $X_t^\theta = X_{\theta t}$. Then
$$\begin{aligned}
\frac{M_\theta}{\sqrt{\theta}} &= f(X_0)\bar{B}_1^\theta + \sqrt{\theta}\int_0^1 \int_0^u g(X_v^\theta)\mathrm{d}\bar{W}_v^\theta\mathrm{d}\bar{B}_u^\theta + \theta\int_0^1 \int_0^u Lf(X_v^\theta)\mathrm{d}v\mathrm{d}\bar{B}_u^\theta, \\
\frac{\langle M\rangle_\theta}{\theta} &= v(X_0) + \sqrt{\theta}\int_0^1 \int_0^u h(X_v^\theta)\mathrm{d}\bar{W}_v^\theta\mathrm{d}u + \theta\int_0^1 \int_0^u Lv(X_v^\theta)\mathrm{d}v\mathrm{d}u.
\end{aligned}$$

It would follow that

$$\frac{\sigma_0(\theta)^2}{\theta} = \frac{E_0[\langle M\rangle_\theta]}{\theta} = v(X_0) + \frac{1}{2}Lv(X_0)\theta + O(\theta^{3/2}),$$

and so

$$\frac{\sigma_0(\theta)}{\sqrt{\theta}} = f(X_0) + \frac{1}{4}\frac{Lv(X_0)}{f(X_0)}\theta + O(\theta^{3/2})$$

under a mild regularity condition. Then, we have (3) with $H = 1/2$, $M^{(0)}_\theta = \bar{B}^\theta_1$ and

$$\begin{aligned}
M^{(1)}_\theta &= \frac{g(X_0)}{f(X_0)}\int_0^1 \bar{W}^\theta_u \mathrm{d}\bar{B}^\theta_u,\\
M^{(2)}_\theta &= -\frac{Lv(X_0)}{4v(X_0)}\bar{B}^\theta_1 + \frac{g'(X_0)c(X_0)}{f(X_0)}\int_0^1\int_0^u \bar{W}^\theta_v \mathrm{d}\bar{W}^\theta_v \mathrm{d}\bar{B}^\theta_u + \frac{Lf(X_0)}{f(X_0)}\int_0^1 u\mathrm{d}\bar{B}^\theta_u,\\
M^{(3)}_\theta &= \frac{h(X_0)}{v(X_0)}\int_0^1 \bar{W}^\theta_u \mathrm{d}u
\end{aligned}$$

again under a mild regularity condition. Further, the derivatives (4) exist in the Schwartz space because $E_0[M^{(i)}_\theta | M^{(0)}_\theta = x]$ and $E_0[|M^{(1)}_\theta|^2 | M^{(0)}_\theta = x]$ are polynomials of x; see e.g., Nualart et al. [NUZ88] or Appendix A below.

3 Asymptotic expansions

3.1 Characteristic function expansion

Here we give an asymptotic expansion of the characteristic function of X_θ. Let

$$Y_\theta = M^{(0)}_\theta + \theta^H M^{(1)}_\theta + \theta^{2H} M^{(2)}_\theta - \frac{\sigma_0(\theta)}{2}\left(1 + \theta^H M^{(3)}_\theta\right).$$

Lemma 1. *For any* $\alpha \in \mathbb{N}\cup\{0\}$,

$$\sup_{|u|\le\theta^{-\epsilon}} |E_0[X^\alpha_\theta e^{iuX_\theta}] - E_0[Y^\alpha_\theta e^{iuY_\theta}]| = o(\theta^{2H+\epsilon}).$$

Proof. Since $|e^{ix} - 1| \le |x|$, we have

$$\begin{aligned}
|E_0[X^\alpha_\theta e^{iuX_\theta}] - E_0[Y^\alpha_\theta e^{iuY_\theta}]| &\le E_0[|X^\alpha_\theta - Y^\alpha_\theta|] + uE_0[|Y_\theta|^\alpha |X_\theta - Y_\theta|]\\
&\le C(\alpha,\epsilon)(1+|u|)\|X_\theta - Y_\theta\|_{1+\epsilon}
\end{aligned}$$

for some constant $C(\alpha,\epsilon) > 0$ by (2). Since $\sigma_0(\theta) = O(\theta^{1/2})$, we obtain the result from (3). □

Lemma 2. *For any* $\delta \in [0,(H-\epsilon)/3)$,

$$\sup_{|u|\le\theta^{-\delta}} \left| E_0[Y^\alpha_\theta e^{iuY_\theta}] - E_0\left[e^{iuM^{(0)}_\theta}\left((M^{(0)}_\theta)^\alpha + A(\alpha,u,M^{(0)}_\theta) + B(\alpha,u,M^{(0)}_\theta)\right)\right]\right| = o(\theta^{2H+\epsilon}),$$

where

$$\begin{aligned}
A_\theta(\alpha,u,x) &= \left(iux^\alpha + \alpha x^{\alpha-1}\right)(E_0[Y_\theta|M_\theta^{(0)}=x]-x),\\
B_\theta(\alpha,u,x) &= \left(-\frac{u^2}{2}x^\alpha + iu\alpha x^{\alpha-1} + \frac{\alpha(\alpha-1)}{2}x^{\alpha-2}\right)\\
&\quad\times\left(\theta^{2H}E_0[|M_\theta^{(1)}|^2|M_\theta^{(0)}=x] - \sigma_0(\theta)\theta^H E_0[M_\theta^{(1)}|M_\theta^{(0)}=x] + \frac{\sigma_0(\theta)^2}{4}\right).
\end{aligned}$$

Proof. This follows from the fact that

$$\left|e^{ix}-1-ix+\frac{x^2}{2}\right| \le \frac{|x|^3}{6}$$

for all $x\in\mathbb{R}$. □

Lemma 3. *Define $q_\theta(x)$ by*

$$\begin{aligned}
q_\theta(x) =&\phi(x) - \theta^H a_\theta^{(1)}(x) - \theta^{2H}a_\theta^{(2)}(x) - \frac{\sigma_0(\theta)}{2}(x\phi(x)-\theta^H a_\theta^{(3)}(x))\\
&+\frac{\theta^{2H}}{2}c_\theta(x) - \frac{\theta^H\sigma_0(\theta)}{2}b_\theta(x) + \frac{\sigma_0(\theta)^2}{8}(x^2-1)\phi(x)
\end{aligned}\tag{5}$$

where $a_\theta^{(i)}$, b_θ and c_θ are defined by (4). Then,

$$\int_{\mathbb{R}} e^{iux}x^\alpha q_\theta(x)\mathrm{d}x = E_0\left[e^{iuM_\theta^{(0)}}\left((M_\theta^{(0)})^\alpha + A(\alpha,u,M_\theta^{(0)}) + B(\alpha,u,M_\theta^{(0)})\right)\right].$$

Proof. Since the density of $M_\theta^{(0)}$ is ϕ by the assumption, this simply follows from integration by parts. □

3.2 Density expansion

Here we derive an asymptotic expansion of the density of X_θ.

Lemma 4. *There exists a density of X_θ under $Q(\cdot|\mathscr{F}_0)$ and for any $\alpha, j\in\mathbb{N}\cup\{0\}$,*

$$\sup_{\theta\in(0,1)}\int |u|^j|E_0[X_\theta^\alpha e^{iuX_\theta}]|\mathrm{d}u < \infty$$

Proof. Note that the distribution of X_θ is Gaussian conditionally on $\mathscr{G}_\theta$ under $Q(\cdot|\mathscr{F}_0)$, with conditional mean

$$-\frac{1}{2\sigma_0(\theta)}\langle M\rangle_\theta + \frac{1}{\sigma_0(\theta)}\int_0^\theta \sqrt{v_t}\rho_t\mathrm{d}W_t$$

and conditional variance

$$\frac{1}{\sigma_0(\theta)^2}\int_0^\theta v_t(1-\rho_t^2)\mathrm{d}t.$$

Therefore, X_θ admits a density $p_\theta(x)$ under $Q(\cdot|\mathscr{F}_0)$. Furthermore, the density function is in the Schwartz space $\mathscr{S}$ and each Schwartz semi-norm is uniformly bounded in θ by (1). Therefore,

$$\begin{aligned}\sup_{\theta\in(0,1)}\int |u|^j|E_0[X_\theta^\alpha e^{iuX_\theta}]|\mathrm{d}u &= \sup_{\theta\in(0,1)}\int\left|\int u^j x^\alpha e^{iux}p_\theta(x)\mathrm{d}x\right|\mathrm{d}u \\ &= \sup_{\theta\in(0,1)}\int\left|\int e^{iux}\partial_x^j(x^\alpha p_\theta(x))\mathrm{d}x\right|\mathrm{d}u<\infty\end{aligned}$$

since the Fourier transform is a continuous linear mapping from $\mathscr{S}$ to $\mathscr{S}$. □

Theorem 1. *Denote by p_θ the density of X_θ under $Q(\cdot|\mathscr{F}_0)$. Then, for any $\alpha\in\mathbb{N}\cup\{0\}$,*

$$\sup_{x\in\mathbb{R}}(1+x^2)^\alpha|p_\theta(x)-q_\theta(x)|=o(\theta^{2H}) \tag{6}$$

as $\theta\to 0$, where q_θ is defined by (5).

Proof. As seen in the proof of Lemma 4, the density p_θ exists in the Schwartz space. By the Fourier identity,

$$(1+x^2)^\alpha(p_\theta(x)-q_\theta(x))=\frac{1}{2\pi}\int\int e^{iuy}(1+y^2)^\alpha(p_\theta(y)-q_\theta(y))\mathrm{d}ye^{-iux}\mathrm{d}u$$

Combining the lemmas in the previous section, taking $\delta\in(0,\min\{\epsilon,(H-\epsilon)/3\})$, we have

$$\int_{|u|\le\theta^{-\delta}}\left|\int e^{iuy}(1+y^2)^\alpha(p_\theta(y)-q_\theta(y))\mathrm{d}y\right|\mathrm{d}u=o(\theta^{2H}).$$

On the other hand,

$$\begin{aligned}\int_{|u|\ge\theta^{-\delta}}\left|\int e^{iuy}(1+y^2)^\alpha p_\theta(y)\mathrm{d}y\right|\mathrm{d}u &\le \theta^{j\delta}\int_{|u|\ge\theta^{-\delta}}|u|^j|E_0[(1+X_\theta^2)^\alpha e^{iuX_\theta}]|\mathrm{d}u \\ &= O(\theta^{j\delta})\end{aligned}$$

for any $j\in\mathbb{N}$ by Lemma 4. The remainder

$$\int_{|u|\ge\theta^{-\delta}}\left|\int e^{iuy}(1+y^2)^\alpha q_\theta(y)\mathrm{d}y\right|\mathrm{d}u$$

is handled in the same manner. □

3.3 Put option price expansion

Here we consider put option prices. Denote by p_θ the density of X_θ as before and consider a normalized put option price

$$\frac{p(Fe^{\sigma_0(\theta)z},\theta)}{F\sigma_0(\theta)}=e^{-r\theta}\int_{-\infty}^{z}\int_{-\infty}^{\zeta}p_\theta(x)\mathrm{d}xe^{\sigma_0(\theta)\zeta}\mathrm{d}\zeta.$$

Lemma 5. *Let $q_\theta(x)$, $\theta > 0$ be a family of functions on $\mathbb{R}$ (not necessarily the one given by (5)). If*

$$\sup_{x\in\mathbb{R}}(1+x^2)^\alpha |p_\theta(x) - q_\theta(x)| = o(\theta^\beta)$$

for some $\alpha > 5/4$ and $\beta > 0$, then for any $z_0 \in \mathbb{R}$,

$$\frac{p(Fe^{\sigma_0(\theta)z},\theta)}{F\sigma_0(\theta)} = e^{-r\theta}\int_{-\infty}^{z}\int_{-\infty}^{\zeta} q_\theta(x)\mathrm{d}x e^{\sigma_0(\theta)\zeta}\mathrm{d}\zeta + o(\theta^\beta)$$

uniformly in $z \le z_0$.

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&e^{-r\theta}\int_{-\infty}^{z}\int_{-\infty}^{\zeta}|p_\theta(x) - q_\theta(x)|\mathrm{d}z e^{\sigma_0(\theta)\zeta}\mathrm{d}\zeta \\
&\le e^{-r\theta}\int_{-\infty}^{z}\sqrt{\int_{-\infty}^{\zeta}\frac{\mathrm{d}x}{(1+x^2)^{2\alpha-1}}}\sqrt{\int_{-\infty}^{\zeta}(1+x^2)^{2\alpha-1}|p_\theta(x)-q_\theta(x)|^2\mathrm{d}z e^{\sigma_0(\theta)\zeta}}\mathrm{d}\zeta \\
&\le \sqrt{\pi}e^{-r\theta+\sigma_0(\theta)z}\sup_{x\in\mathbb{R}}(1+x^2)^\alpha|p_\theta(x)-q_\theta(x)|\int_{-\infty}^{z}\sqrt{\int_{-\infty}^{\zeta}\frac{\mathrm{d}x}{(1+x^2)^{2\alpha-1}}}\mathrm{d}\zeta,
\end{aligned}$$

which is $o(\theta^\beta)$ if $\alpha > 5/4$. □

Proposition 1. *Suppose we have (6) with q_θ of the form*

$$\begin{aligned}
q_\theta(x) =&\phi\left(x+\frac{\sigma_0(\theta)}{2}\right)\left\{1+\kappa_3(\theta)\left(H_3\left(x+\frac{\sigma_0(\theta)}{2}\right)-\sigma_0(\theta)H_2\left(x+\frac{\sigma_0(\theta)}{2}\right)\right)\theta^H\right\} \\
&+\phi(x)\left(\kappa_4(\theta)H_4(x)+\frac{\kappa_3(\theta)^2}{2}H_6(x)\right)\theta^{2H}
\end{aligned} \tag{7}$$

with bounded functions $\kappa_3(\theta)$ and $\kappa_4(\theta)$ of θ, where H_k is the kth Hermite polynomial:

$$H_1(x) = x,\quad H_2(x) = x^2-1,\quad H_3(x) = x^3-3x,\quad H_4(x) = x^4-6x^2+3,\dots$$

Then, for any $z_0 \in \mathbb{R}$,

$$\begin{aligned}
\frac{p(Fe^{\sigma_0(\theta)z},\theta)}{Fe^{-r\theta}\sigma_0(\theta)} = &\frac{1}{\sigma_0(\theta)}\left(\Phi\left(z+\frac{\sigma_0(\theta)}{2}\right)e^{\sigma_0(\theta)z}-\Phi\left(z-\frac{\sigma_0(\theta)}{2}\right)\right) \\
&+\kappa_3(\theta)\phi\left(z+\frac{\sigma_0(\theta)}{2}\right)H_1\left(z+\frac{\sigma_0(\theta)}{2}\right)e^{\sigma_0(\theta)z}\theta^H \\
&+\phi(z)\left(\kappa_4(\theta)H_2(z)+\frac{\kappa_3(\theta)^2}{2}H_4(z)\right)\theta^{2H}+o(\theta^{2H})
\end{aligned}$$

uniformly in $z \le z_0$.

Proof. This is a direct consequence of the previous lemma. For example,

$$\frac{\mathrm{d}}{\mathrm{d}z}\left\{e^{-\sigma_0(\theta)z}\frac{\mathrm{d}}{\mathrm{d}z}\left\{\frac{1}{\sigma_0(\theta)}\left(\Phi\left(z+\frac{\sigma_0(\theta)}{2}\right)e^{\sigma_0(\theta)z}-\Phi\left(z-\frac{\sigma_0(\theta)}{2}\right)\right)\right\}\right\} = \phi\left(z+\frac{\sigma_0(\theta)}{2}\right).$$

The derivative of $H_k(z)\phi(z)$ is $-H_{k+1}(z)\phi(z)$. Recall also $\sigma_0(\theta) = O(\sqrt{\theta})$. □

3.4 Implied volatility expansion

Here we give an expansion formula for the Black-Scholes implied volatility. Denote by $p_{\mathrm{BS}}(K,\theta,\sigma)$ the put option price with strike price K and maturity θ under the Black-Scholes model with volatility parameter $\sigma > 0$. Given a put option price $p(K,\theta)$, $K = Fe^k$, the implied volatility $\sigma_{\mathrm{BS}}(k,\theta)$ is defined through

$$p_{\mathrm{BS}}(K,\theta,\sigma_{\mathrm{BS}}(k,\theta)) = p(K,\theta).$$

Theorem 2. *Suppose we have (6) with q_θ of the form (7). Then, for any $z \in \mathbb{R}$,*

$$\begin{aligned}&\sigma_{BS}(\sqrt{\theta}z,\theta)\\ &= \kappa_2\left\{1+\kappa_3\left(\frac{z}{\kappa_2}+\frac{\kappa_2\sqrt{\theta}}{2}\right)\theta^H+\left(\frac{3\kappa_3^2}{2}-\kappa_4+(\kappa_4-3\kappa_3^2)\frac{z^2}{\kappa_2^2}\right)\theta^{2H}\right\}+o(\theta^{2H}),\end{aligned}$$

where $\kappa_2 = \kappa_2(\theta) = \sigma_0(\theta)/\sqrt{\theta}$, $\kappa_3 = \kappa_3(\theta)$ and $\kappa_4 = \kappa_4(\theta)$.

Proof. Step 1). Fix $z \in \mathbb{R}$. Note that

$$P_\theta(\sigma) := \frac{p_{\mathrm{BS}}(Fe^{\sqrt{\theta}z},\theta,\sigma)}{Fe^{-r\theta}\sqrt{\theta}} = \frac{1}{\sqrt{\theta}}\left(\Phi\left(\frac{z}{\sigma}+\frac{\sigma\sqrt{\theta}}{2}\right)e^{\sqrt{\theta}z}-\Phi\left(\frac{z}{\sigma}-\frac{\sigma\sqrt{\theta}}{2}\right)\right) \tag{8}$$

and that

$$P_\theta : [0,\infty] \to \left[\frac{(e^{\sqrt{\theta}z}-1)_+}{\sqrt{\theta}}, \frac{e^{\sqrt{\theta}z}}{\sqrt{\theta}}\right]$$

is a strictly increasing function. From (8) and Proposition 1, we have

$$\begin{aligned}\frac{p(Fe^{\sqrt{\theta}z},\theta)}{Fe^{-r\theta}\sqrt{\theta}} =& P_\theta(\kappa_2)+\kappa_2\kappa_3\phi\left(\frac{z}{\kappa_2}+\frac{\kappa_2\sqrt{\theta}}{2}\right)H_1\left(\frac{z}{\kappa_2}+\frac{\kappa_2\sqrt{\theta}}{2}\right)e^{\sqrt{\theta}z}\theta^H\\ &+\kappa_2\phi\left(\frac{z}{\kappa_2}\right)\left(\kappa_4H_2\left(\frac{z}{\kappa_2}\right)+\frac{\kappa_3^2}{2}H_4\left(\frac{z}{\kappa_2}\right)\right)\theta^{2H}+o(\theta^{2H})\\ =& P_\theta(\kappa_2)+O(\theta^H).\end{aligned}$$

Therefore

$$\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) = P_\theta^{-1}(P_\theta(\kappa_2)+O(\theta^H)).$$

By (1), κ_2 is bounded in θ, say, by $L > 0$. The function P_θ converges as $\theta \to 0$ to

$$P_0(\sigma) := z\Phi\left(\frac{z}{\sigma}\right)+\sigma\phi\left(\frac{z}{\sigma}\right)$$

pointwize, and by Dini's theorem, this convergence is uniform on $[0,L]$. Since the limit function P_0 is strictly increasing, the inverse functions P_θ^{-1} converges to P_0^{-1}. Again by Dini's theorem, this convergence is uniform and in particular, P_θ^{-1} are equicontinuous. Thus we conclude $\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)-\kappa_2 \to 0$ as $\theta \to 0$. Then, write $\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) = \kappa_2+\beta(\theta)$ and substitute this to the equation $P_\theta(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)) = P_\theta(\kappa_2) + O(\theta^H)$. The Taylor expansion gives $\beta(\theta) = O(\theta^H)$.

Step 2). From (8) we have

$$P_\theta(\sigma) = \sigma F_1\left(\frac{z}{\sigma}\right) + \frac{\sigma^2\sqrt{\theta}}{2}F_2\left(\frac{z}{\sigma}\right) + \frac{\sigma^3\theta}{6}F_3\left(\frac{z}{\sigma}\right) + o(\theta),$$

where

$$F_1(x) = x\Phi(x) + \phi(x), \quad F_2(x) = x^2\Phi(x) + x\phi(x), \quad F_3(x) = x^3\Phi(x) + \left(x^2 - \frac{1}{4}\right)\phi(x).$$

Using that

$$\partial_\sigma\left\{\sigma F_1\left(\frac{z}{\sigma}\right)\right\} = \phi\left(\frac{z}{\sigma}\right),$$

we have

$$\begin{aligned}
&\kappa_2 F_1\left(\frac{z}{\kappa_2}\right) + \frac{\kappa_2^2\sqrt{\theta}}{2}F_2\left(\frac{z}{\kappa_2}\right) + \kappa_2\phi\left(\frac{z}{\kappa_2}\right)\kappa_3 H_1\left(\frac{z}{\kappa_2}\right)e^{\sqrt{\theta}z}\theta^H \\
&= \sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)F_1\left(\frac{z}{\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)}\right) + \frac{\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)^2\sqrt{\theta}}{2}F_2\left(\frac{z}{\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta)}\right) + O(\theta^{2H}) \\
&= \kappa_2 F_1\left(\frac{z}{\kappa_2}\right) + \frac{\kappa_2^2\sqrt{\theta}}{2}F_2\left(\frac{z}{\kappa_2}\right) + \phi\left(\frac{z}{\kappa_2}\right)(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2) + O(\theta^{2H}),
\end{aligned}$$

from which we conclude $\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) = \kappa_2 + \kappa_3 z e^{\sqrt{\theta}z}\theta^H + O(\theta^{2H})$.

Step 3). Using that

$$\partial_\sigma^2\left\{\sigma F_1\left(\frac{z}{\sigma}\right)\right\} = \frac{z^2}{\sigma^3}\phi\left(\frac{z}{\sigma}\right), \quad \partial_\sigma\left\{\sigma^2 F_2\left(\frac{z}{\sigma}\right)\right\} = z\phi\left(\frac{z}{\sigma}\right),$$

we obtain

$$\begin{aligned}
&\kappa_2\phi\left(\frac{z}{\kappa_2} + \frac{\kappa_2\sqrt{\theta}}{2}\right)\left(\kappa_3 H_1\left(\frac{z}{\kappa_2} + \frac{\kappa_2\sqrt{\theta}}{2}\right)e^{\sqrt{\theta}z}\theta^H + \left(\kappa_4 H_2\left(\frac{z}{\kappa_2}\right) + \frac{\kappa_3^2}{2}H_4\left(\frac{z}{\kappa_2}\right)\right)\theta^{2H}\right) \\
&= \frac{p(Fe^{\sqrt{\theta}z},\theta)}{Fe^{-r\theta}\sqrt{\theta}} - P_\theta(\kappa_2) + o(\theta^{2H}) \\
&= P_\theta(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,z)) - P_\theta(\kappa_2) + o(\theta^{2H}) \\
&= \partial_\sigma\left\{\sigma F_1\left(\frac{z}{\sigma}\right)\right\}|_{\sigma=\kappa_2}(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2) + \frac{1}{2}\partial_\sigma^2\left\{\sigma F_1\left(\frac{z}{\sigma}\right)\right\}|_{\sigma=\kappa_2}(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2)^2 \\
&\quad + \frac{\sqrt{\theta}}{2}\partial_\sigma\left\{\sigma^2 F_2\left(\frac{z}{\sigma}\right)\right\}\Big|_{\sigma=\kappa_2}(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2) + o(\theta^{2H}) \\
&= \phi\left(\frac{z}{\kappa_2}\right)(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2) + \frac{\sqrt{\theta}}{2}z\phi\left(\frac{z}{\kappa_2}\right)(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2) \\
&\quad + \frac{z^2}{2\kappa_2^3}\phi\left(\frac{z}{\kappa_2}\right)(\sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2)^2 + o(\theta^{2H})
\end{aligned}$$

from Proposition 1 and Step 2. The left hand side is further expanded as

$$\kappa_2\phi\left(\frac{z}{\kappa_2}\right)\left\{\kappa_3 H_1\left(\frac{z}{\kappa_2}\right)e^{\sqrt{\theta}z}\theta^H - \kappa_3 H_2\left(\frac{z}{\kappa_2}\right)\frac{\kappa_2}{2}\theta^{H+1/2}\right.$$
$$\left. + \left(\kappa_4 H_2\left(\frac{z}{\kappa_2}\right) + \frac{\kappa_3^2}{2}H_4\left(\frac{z}{\kappa_2}\right)\right)\theta^{2H}\right\} + o(\theta^{2H}).$$

Denote $\gamma(\theta) = \sigma_{\mathrm{BS}}(\sqrt{\theta}z,\theta) - \kappa_2 - \kappa_3 z e^{\sqrt{\theta}z}\theta^H$ and substitute this to obtain

$$\begin{aligned}
\gamma(\theta) &= -\kappa_3 H_2\left(\frac{z}{\kappa_2}\right)\frac{\kappa_2^2}{2}\theta^{H+1/2} + \kappa_2\left(\kappa_4 H_2\left(\frac{z}{\kappa_2}\right) + \frac{\kappa_3^2}{2}H_4\left(\frac{z}{\kappa_2}\right)\right)\theta^{2H} \\
&\quad - \frac{\kappa_3}{2}z^2\theta^{H+1/2} - \frac{\kappa_3^2}{2\kappa_2^3}z^4\theta^{2H} + o(\theta^{2H}) \\
&= \left(\frac{\kappa_2^2}{2} - z^2\right)\kappa_3\theta^{H+1/2} + \kappa_2\left((\kappa_4 - 3\kappa_3^2)\frac{z^2}{\kappa_2^2} + \frac{3}{2}\kappa_3^2 - \kappa_4\right)\theta^{2H} + o(\theta^{2H}),
\end{aligned}$$

from which we conclude the result. □

4 Asymptotics for at-the-money skew and curvature

Here we derive the asymptotic behavior of at-the-money implied volatility skew and curvature. They are defined respectively as the first and the second derivatives of the implied volatility at $k = 0$. The skew behavior is especially important in order to argue the consistency of a model to the empirically observed power law.

Theorem 3. *Suppose we have (6) with q_θ of the form (7). Then,*

$$\partial_k\sigma_{BS}(0,\theta) = \kappa_3(\theta)\theta^{H-1/2} + o(\theta^{2H-1/2}),$$
$$\partial_k^2\sigma_{BS}(0,\theta) = 2\frac{\kappa_4(\theta) - 3\kappa_3(\theta)^2}{\kappa_2(\theta)}\theta^{2H-1} + o(\theta^{2H-1}).$$

Proof. It is known (see e.g., Fukasawa [Fuk12]) that

$$\begin{aligned}
\partial_k\sigma_{\mathrm{BS}}(k,\theta) &= \frac{Q(k \geq \sigma_0(\theta)X_\theta|\mathscr{F}_0) - \Phi(f_2(k,\theta))}{\sqrt{\theta}\phi(f_2(k,\theta))}, \\
\partial_k^2\sigma_{\mathrm{BS}}(k,\theta) &= \frac{p_\theta(k/\sigma_0(\theta))}{\sigma_0(\theta)\sqrt{\theta}\phi(f_2(k,\theta))} - \sigma_{\mathrm{BS}}(k,\theta)\partial_k f_1(k,\theta)\partial_k f_2(k,\theta),
\end{aligned} \tag{9}$$

where

$$f_1(k,\theta) = \frac{k}{\sqrt{\theta}\sigma_{\mathrm{BS}}(k,\theta)} - \frac{\sqrt{\theta}\sigma_{\mathrm{BS}}(k,\theta)}{2}, \quad f_2(k,\theta) = \frac{k}{\sqrt{\theta}\sigma_{\mathrm{BS}}(k,\theta)} + \frac{\sqrt{\theta}\sigma_{\mathrm{BS}}(k,\theta)}{2}.$$

Since the condition of Proposition 1 is met, we have

$$Q(0 \geq X_\theta|\mathscr{F}_0) = \Phi\left(\frac{\sigma_0(\theta)}{2}\right) + \kappa_3(\theta)\phi\left(\frac{\sigma_0(\theta)}{2}\right)\theta^H + o(\theta^{2H}).$$

On the other hand, by Theorem 2,

$$f_2(0,\theta) = \frac{\sqrt{\theta}}{2}\kappa_2(\theta) + O(\theta^{2H+1/2})$$

and so,

$$\begin{aligned}\Phi(f_2(0,\theta)) &= \Phi\left(\frac{\sigma_0(\theta)}{2}\right) + O(\theta^{2H+1/2}),\\ \phi(f_2(0,\theta)) &= \phi(0) - \phi(0)\frac{\theta}{8}\kappa_2(\theta)^2 + O(\theta^{2H+1}).\end{aligned}$$

Then, it follows from (9) that

$$\partial_k \sigma_{\mathrm{BS}}(0,\theta) = \kappa_3(\theta)\theta^{H-1/2} + o(\theta^{2H-1/2}). \tag{10}$$

Further, under the condition, we have

$$p_\theta(0) = \phi\left(\frac{\sigma_0(\theta)}{2}\right)\left\{1 - \frac{\kappa_3(\theta)}{2}\sigma_0(\theta)\theta^H + \left(3\kappa_4(\theta) - 15\frac{\kappa_3(\theta)^2}{2}\right)\theta^{2H}\right\} + o(\theta^{2H}).$$

On the other hand, by Theorem 2 and (10),

$$\begin{aligned}&\sigma_{\mathrm{BS}}(0,\theta)\partial_k f_1(0,\theta)\partial_k f_2(0,\theta)\\ &=\frac{1}{\sigma_{\mathrm{BS}}(0,\theta)\theta} + O(\theta^{2H})\\ &=\frac{1}{\kappa_2(\theta)\theta}\left(1 - \frac{1}{2}\kappa_2(\theta)\kappa_3(\theta)\theta^{H+1/2} - \left(\frac{3}{2}\kappa_3(\theta)^2 - \kappa_4(\theta)\right)\theta^{2H}\right) + o(\theta^{2H-1}).\end{aligned}$$

Then, it follows from (9) that

$$\partial_k^2 \sigma_{\mathrm{BS}}(0,\theta) = \frac{2\kappa_4(\theta) - 6\kappa_3(\theta)^2}{\kappa_2(\theta)}\theta^{2H-1} + o(\theta^{2H-1}),$$

which completes the proof. □

5 The rough Bergomi model

Here we show that the rough Bergomi model proposed by [BFG16] fits the framework and compute the expansion terms. Let $\rho_t = \rho \in (-1,1)$ be a constant and

$$\mathrm{d}\log v_t = \eta \mathrm{d}W_t^H,$$

where $\eta > 0$ is a constant and W^H is a fractional Brownian motion with the Hurst parameter $H \in (0,1/2)$, given as

$$W_t^H = c_H\left\{\int_0^t (t-s)^{H-1/2}\mathrm{d}W_s + \int_{-\infty}^0 (t-s)^{H-1/2} - (-s)^{H-1/2}\mathrm{d}W_s\right\}$$

with a normalizing constant $c_H > 0$. Since v_t is log-normally distributed, (1) holds by Jensen's inequality. We have

$$v_t = v^0(t)\exp\left\{\eta_H\sqrt{2H}\int_0^t (t-s)^{H-1/2}\mathrm{d}W_s - \frac{\eta_H^2}{2}t^{2H}\right\},$$

where $\eta_H = \eta c_H/\sqrt{2H}$. Now we state the main result of this section.

Theorem 4. *We have (6) for q_θ given by (7) with*

$$\kappa_3(\theta) = \rho\eta_H\sqrt{\frac{H}{2}}\frac{1}{\theta^H\sigma_0(\theta)^3}\int_0^\theta \exp\left\{-\frac{\eta_H^2}{8}t^{2H}\right\}\int_0^t (t-s)^{H-1/2}\sqrt{v^0(s)}\mathrm{d}s\, v^0(t)\mathrm{d}t,$$
$$\kappa_4(\theta) = \frac{(1+2\rho^2)\eta_H^2 H}{(2H+1)^2(2H+2)} + \frac{\rho^2\eta_H^2 H\beta(H+3/2, H+3/2)}{2(H+1/2)^2},$$

where β is the beta function.

Proof. The conditions (2) and (3) follow from Lemma 6 below. The functions $a_\theta^{(i)}$ and c_θ are computed in Lemmas 7, 8, 9 and 10 below. The function b_θ is obtained as the the derivative of $a_\theta^{(1)}$. They are apparently rapidly decreasing smooth functions. Then, by Theorem 1, it suffices to show that q_θ defined by (5) has the form (7) up to $o(\theta^{2H})$ with $\kappa_3(\theta)$ and $\kappa_4(\theta)$ specified above.

By the Taylor expansion, using that the derivative of $a_\theta^{(1)}$ is b_θ and that $\sigma_0(\theta) = O(\sqrt{\theta})$, it is easy to verify

$$\begin{aligned} q_\theta(x) =& \phi\left(x + \frac{\sigma_0(\theta)}{2}\right) - \theta^H a_\theta^{(1)}\left(x + \frac{\sigma_0(\theta)}{2}\right) - \theta^{2H}a_\theta^{(2)}(x) + \frac{\theta^{2H}}{2}c_\theta(x) \\ &+ \frac{\sigma_0(\theta)\theta^H}{2}a_\theta^{(3)}\left(x + \frac{\sigma_0(\theta)}{2}\right) + O(\theta^{1+H}) \end{aligned}$$

in the Schwarz space. The rest is straightforward. □

Proposition 1 and Theorems 2 and 3 are therefore valid here. The resulting formula of the implied volatility expansion turns out reduce the Bergomi-Guyon expansion formula formally derived in [BFG16] when $H < 1/2$ and the forward variance curve is flat, that is, when v^0 is constant. Note however that there is a typo in the second order term in [BFG16]. Numerical experiments are given in that paper. When v^0 is constant, the same formula can be formally obtained also by expanding the rate function appeared in the large deviation result of [FZ17]; see [BFG+17] for a rigorous treatment in this approach.

In order to prove Lemmas below, we need some preparation. Let H_k, $k = 0, 1, \dots$ be the Hermite polynomials as before:

$$H_k(x) = (-1)^k e^{x^2/2}\frac{\mathrm{d}^k}{\mathrm{d}x^k}e^{-x^2/2}$$

and $H_k(x,a) = a^{k/2} H_k(x/\sqrt{a})$ for $a > 0$. As is well-known, we have

$$\exp\left\{ux - \frac{au^2}{2}\right\} = \sum_{k=0}^{\infty} H_k(x,a) \frac{u^k}{k!}$$

and for any continuous local martingale M and $n \in \mathbb{N}$,

$$\mathrm{d}L_t^{(n)} = nL_t^{(n-1)} \mathrm{d}M_t, \tag{11}$$

where $L^{(k)} = H_k(M, \langle M \rangle)$ for $k \in \mathbb{N}$. See, e.g., Revuz and Yor [RY13].

Define $\hat{W}$, $\hat{W}'$, $\hat{B}$ by

$$\hat{W}_t = \frac{1}{\sigma_0(\theta)} \int_0^{\tau^{-1}(t)} \sqrt{v^0(s)} \mathrm{d}W_s, \quad \hat{W}'_t = \frac{1}{\sigma_0(\theta)} \int_0^{\tau^{-1}(t)} \sqrt{v^0(s)} \mathrm{d}W'_s$$

and $\hat{B} = \rho \hat{W} + \sqrt{1-\rho^2} \hat{W}'$, where

$$\tau(s) = \frac{1}{\sigma_0(\theta)^2} \int_0^s v^0(t) \mathrm{d}t.$$

Then, $(\hat{W}, \hat{W}')$ is a 2-dimensional Brownian motion under E_0 and for any square-integrable function f,

$$\int_0^a f(s) \mathrm{d}W_s = \sigma_0(\theta) \int_0^{\tau(a)} \frac{f(\tau^{-1}(t))}{\sqrt{v^0(\tau^{-1}(t))}} \mathrm{d}\hat{W}_t.$$

Therefore,

$$M_\theta = \sigma_0(\theta) \int_0^1 \exp\left\{\theta^H F_t^t - \frac{\eta_H^2}{4} |\tau^{-1}(t)|^{2H}\right\} \mathrm{d}\hat{B}_t$$

where

$$F_u^t = \eta_H \sqrt{\frac{H}{2}} \frac{\sigma_0(\theta)}{\theta^H} \int_0^u \frac{(\tau^{-1}(t) - \tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}} \mathrm{d}\hat{W}_s, \quad u \in [0,t].$$

Let

$$G_t^{(k)} = H_k(F_t^t, \langle F^t \rangle_t).$$

Then, we have

$$\begin{aligned} M_\theta &= \sigma_0(\theta) \int_0^1 \exp\left\{-\frac{\eta_H^2}{8} |\tau^{-1}(t)|^{2H}\right\} \exp\left\{\theta^H F_t^t - \frac{\theta^{2H}}{2} \langle F^t \rangle_t\right\} \mathrm{d}\hat{B}_t \\ &= \sigma_0(\theta) \int_0^1 \exp\left\{-\frac{\eta_H^2}{8} |\tau^{-1}(t)|^{2H}\right\} \sum_{k=0}^{\infty} G_t^{(k)} \frac{\theta^{Hk}}{k!} \mathrm{d}\hat{B}_t. \end{aligned}$$

Lemma 6. *We have (3) with*

$$\begin{aligned}
M_\theta^{(0)} &= \hat{B}_1,\\
M_\theta^{(1)} &= \int_0^1 h_\theta(t) G_t^{(1)} \mathrm{d}\hat{B}_t,\\
M_\theta^{(2)} &= \int_0^1 \left\{ \frac{h_\theta(t)-1}{\theta^{2H}} + h_\theta(t)\frac{G_t^{(2)}}{2} \right\} \mathrm{d}\hat{B}_t,\\
M_\theta^{(3)} &= 2\int_0^1 F_t^t \mathrm{d}t,
\end{aligned}$$

where

$$h_\theta(t) = \exp\left\{ -\frac{\eta_H^2}{8} |\tau^{-1}(t)|^{2H} \right\}.$$

Proof. For $M_\theta^{(i)}$, $i = 0,1,2$, it suffices to show

$$\left\| \int_0^1 h_\theta(t) \sum_{k=J}^\infty G_t^{(k)} \frac{\theta^{Hk}}{k!} \mathrm{d}\hat{B}_t \right\|_2 = O(\theta^{HJ})$$

for any $J \geq 3$. The proof for $M_\theta^{(3)}$ is similar and so omitted. It suffices to show

$$E_0\left[\int_0^1 \left| \sum_{k=J}^\infty G_t^{(k)} \frac{\theta^{Hk}}{k!} \right|^2 \mathrm{d}t \right] = O(\theta^{2HJ}).$$

By the Cauchy-Schwarz inequality, the left hand side is dominated by

$$\sum_{k=J}^\infty \theta^{Hk} \sum_{k=J}^\infty \frac{\theta^{Hk}}{(k!)^2} \int_0^1 E_0[|G_t^{(k)}|^2] \mathrm{d}t$$

Let

$$G_{t,s}^{(k)} = H_k(F_s^t, \langle F^t \rangle_s), \quad s \in [0,t].$$

Then, by (11),

$$\begin{aligned}
E_0[|G_t^{(k)}|^2] &= E_0[|G_{t,t}^{(k)}|^2]\\
&= k^2 \int_0^t E_0[|G_{t,s}^{(k-1)}|^2] \mathrm{d}\langle F^t \rangle_s\\
&= k^2(k-1)^2 \int_0^t \int_0^{s_1} E_0[|G_{t,s_2}^{(k-2)}|^2] \mathrm{d}\langle F^t \rangle_{s_2} \mathrm{d}\langle F^t \rangle_{s_1}\\
&\leq (k!)^2 \langle F^t \rangle_t^k = (k!)^2 \left(\frac{\eta_H^2}{4} \frac{|\tau^{-1}(t)|^{2H}}{\theta^{2H}} \right)^k.
\end{aligned}$$

Note that $\tau^{-1}(t) \leq \tau^{-1}(1) = \theta$. Therefore, for sufficiently small θ,

$$\sum_{k=J}^\infty \theta^{Hk} \sum_{k=J}^\infty \frac{\theta^{Hk}}{(k!)^2} \int_0^1 E_0[|G_t^{(k)}|^2] \mathrm{d}t \leq \left(\frac{\eta_H^2}{4} \right)^J \frac{\theta^{2HJ}}{(1-\theta^H)(1-\theta^H \eta_H^2/4)},$$

which completes the proof. □

Now we compute $a_\theta^{(i)}$, b_θ and c_θ based on Lemma 6. The following lemmas follow from the results in Section VII.A by straightforward computations.

Lemma 7.

$$\begin{aligned}
a_\theta^{(1)}(x) &= -H_3(x)\phi(x)\rho\eta_H\sqrt{\frac{H}{2}}\frac{\sigma_0(\theta)}{\theta^H}\int_0^1 h_\theta(t)\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s\mathrm{d}t\\
&= -H_3(x)\phi(x)\rho\eta_H\sqrt{\frac{H}{2}}\\
&\quad\times\frac{1}{\theta^H\sigma_0(\theta)^3}\int_0^\theta \exp\left\{-\frac{\eta_H^2}{8}t^{2H}\right\}\int_0^t (t-s)^{H-1/2}\sqrt{v^0(s)}\mathrm{d}s v^0(t)\mathrm{d}t\\
&\sim -H_3(x)\phi(x)\frac{\rho\eta_H\sqrt{2H}}{2(H+1/2)(H+3/2)}.
\end{aligned}$$

Lemma 8.

$$\begin{aligned}
a_\theta^{(2)}(x) &= -H_2(x)\phi(x)\int_0^1 \frac{h_\theta(t)-1}{\theta^{2H}}\mathrm{d}t\\
&- H_4(x)\phi(x)\rho^2\frac{\eta_H^2 H}{4}\frac{\sigma_0(\theta)^2}{\theta^{2H}}\int_0^1 h_\theta(t)\left(\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s\right)^2\mathrm{d}t\\
&\sim H_2(x)\phi(x)\int_0^1 \frac{\eta_H^2}{8\theta^{2H}}|\tau^{-1}(t)|^{2H}\mathrm{d}t - H_4(x)\phi(x)\rho^2\frac{\eta_H^2 H}{(2H+1)^2(2H+2)}.
\end{aligned}$$

Lemma 9.

$$\begin{aligned}
a_\theta^{(3)}(x) &= -2H_2(x)\phi(x)\rho\eta_H\sqrt{\frac{H}{2}}\frac{\sigma_0(\theta)}{\theta^H}\int_0^1\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s\mathrm{d}t\\
&\sim -2H_2(x)\phi(x)\frac{\rho\eta_H\sqrt{2H}}{2(H+1/2)(H+3/2)}.
\end{aligned}$$

Lemma 10.

$$
\begin{aligned}
c_\theta(x) = {} & H_6(x)\phi(x)\rho^2\frac{\eta_H^2 H}{2}\frac{\sigma_0(\theta)^2}{\theta^{2H}}\left(\int_0^1 h_\theta(t)\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s\mathrm{d}t\right)^2 \\
& + H_4(x)\phi(x)\rho^2\frac{\eta_H^2 H}{2}\frac{\sigma_0(\theta)^2}{\theta^{2H}}\int_0^1 h_\theta(t)^2\left(\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s\right)^2\mathrm{d}t \\
& + H_4(x)\phi(x)\rho^2\eta_H^2 H\frac{\sigma_0(\theta)^2}{\theta^{2H}}\int_0^1 h_\theta(t)\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}s \\
& \qquad \times\int_t^1 h_\theta(u)\frac{(\tau^{-1}(u)-\tau^{-1}(t))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(t))}}\mathrm{d}u\mathrm{d}t \\
& + H_4(x)\phi(x)\frac{\eta_H^2 H}{2}\frac{\sigma_0(\theta)^2}{\theta^{2H}}\int_0^1 h_\theta(t)^2\left(\int_s^1 \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{H-1/2}}{\sqrt{v^0(\tau^{-1}(s))}}\mathrm{d}t\right)^2\mathrm{d}s \\
& + H_2(x)\phi(x)\frac{\eta_H^2 H}{2}\frac{\sigma_0(\theta)^2}{\theta^{2H}}\int_0^1 h_\theta(t)^2\int_0^t \frac{(\tau^{-1}(t)-\tau^{-1}(s))^{2H-1}}{v^0(\tau^{-1}(s))}\mathrm{d}s\mathrm{d}t \\
\sim {} & H_6(x)\phi(x)\rho^2\frac{\eta_H^2 H}{2(H+1/2)^2(H+3/2)^2} + H_4(x)\phi(x)\frac{2(1+\rho^2)\eta_H^2 H}{(2H+1)^2(2H+2)} \\
& + H_4(x)\phi(x)\frac{\rho^2\eta_H^2 H\beta(H+3/2,H+3/2)}{(H+1/2)^2} \\
& + H_2(x)\phi(x)\int_0^1 \frac{\eta_H^2}{4\theta^{2H}}|\tau^{-1}(t)|^{2H}\mathrm{d}t.
\end{aligned}
$$

VII.A Conditional expectations of Wiener-Itô integrals

Here we collect results on the conditional expectations of Wiener-Itô integrals that follow from Proposition 3 of Nualart et al [NUZ88]. Let $x \in \mathbb{R}$ and B be a standard Brownian motion ($B_0 = 0$). Let f be a continuous function on

$$\left\{(s,t) \in (0,1)^2; s < t\right\}$$

with

$$\int_0^1\int_0^t |f(s,t)|^2\mathrm{d}s\mathrm{d}t < \infty.$$

Lemma 11.

$$
\begin{aligned}
E\left[\int_0^1\int_0^t f(s,t)\mathrm{d}B_s\mathrm{d}t \;\Big|\; B_1=x\right] &= H_1(x)\int_0^1\int_0^t f(s,t)\mathrm{d}s\mathrm{d}t,\\
E\left[\int_0^1\int_0^t f(s,t)\mathrm{d}B_s\mathrm{d}B_t \;\Big|\; B_1=x\right] &= H_2(x)\int_0^1\int_0^t f(s,t)\mathrm{d}s\mathrm{d}t,\\
E\left[\int_0^1\left(\int_0^t f(s,t)\mathrm{d}B_s\right)^2\mathrm{d}B_t \;\Big|\; B_1=x\right] &= H_3(x)\int_0^1\left(\int_0^t f(s,t)\mathrm{d}s\right)^2\mathrm{d}t\\
&\quad + H_1(x)\int_0^1\int_0^t f(s,t)^2\mathrm{d}s\mathrm{d}t,\\
E\left[\int_0^1\left(\int_s^1 f(s,t)\mathrm{d}B_t\right)^2\mathrm{d}s \;\Big|\; B_1=x\right] &= H_2(x)\int_0^1\left(\int_s^1 f(s,t)\mathrm{d}t\right)^2\mathrm{d}s\\
&\quad + \int_0^1\int_s^1 f(s,t)^2\mathrm{d}t\mathrm{d}s.
\end{aligned}
$$

Lemma 12.

$$
\begin{aligned}
&E\left[\left(\int_0^1\int_0^t f(s,t)\mathrm{d}B_s\mathrm{d}B_t\right)^2 \;\Big|\; B_1=x\right] - \int_0^1\int_0^t f(s,t)^2\mathrm{d}s\mathrm{d}t\\
&= H_4(x)\left(\int_0^1\int_0^t f(s,t)\mathrm{d}s\mathrm{d}t\right)^2 + H_2(x)\int_0^1\left(\int_0^t f(s,t)\mathrm{d}s + \int_t^1 f(t,u)\mathrm{d}u\right)^2\mathrm{d}t.
\end{aligned}
$$

Part VI

Markovian approximation of rough volatility models

CHAPTER VIII

Multi-factor approximation of rough volatility models

Abstract

Rough volatility models are very appealing because of their remarkable fit of both historical and implied volatilities. However, due to the non-Markovian and non-semimartingale nature of the volatility process, there is no simple way to simulate efficiently such models, which makes risk management of derivatives an intricate task. In this paper, we design tractable multi-factor stochastic volatility models approximating rough volatility models and enjoying a Markovian structure. Furthermore, we apply our procedure to the specific case of the rough Heston model. This in turn enables us to derive a numerical method for solving fractional Riccati equations appearing in the characteristic function of the log-price in this setting.

Keywords: Rough volatility models, rough Heston models, stochastic Volterra equations, affine Volterra processes, fractional Riccati equations, limit theorems.

1 Introduction

Empirical studies of a very wide range of assets volatility time-series in [GJR18] have shown that the dynamics of the log-volatility are close to that of a fractional Brownian motion W^H with a small Hurst parameter H of order 0.1. Recall that a fractional Brownian motion W^H can be built from a two-sided Brownian motion thanks to the Mandelbrot-van Ness representation

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

The fractional kernel $(t-s)^{H-\frac{1}{2}}$ is behind the $H-\varepsilon$ Hölder regularity of the volatility for any $\varepsilon > 0$. For small values of the Hurst parameter H, as observed empirically, stochastic volatility models involving the fractional kernel are called rough volatility models.

Aside from modeling historical volatility dynamics, rough volatility models reproduce accurately with very few parameters the behavior of the implied volatility surface, see [BFG16] and Chapter V, especially the at-the-money skew, see [Fuk11]. Moreover, microstructural foundations of rough volatility are studied in [JR16b] and Chapter I.

In this paper, we are interested in a class of rough volatility models where the dynamics of the asset price S and its stochastic variance V are given by

$$dS_t = S_t\sqrt{V_t}dW_t, \quad S_0 > 0, \tag{1}$$

$$V_t = V_0 + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-u)^{H-\frac{1}{2}}(\theta(u) - \lambda V_u)du + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-u)^{H-\frac{1}{2}}\sigma(V_u)dB_u, \tag{2}$$

for all $t \in [0,T]$, on some filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$. Here T is a positive time horizon, the parameters λ and V_0 are non-negative, $H \in (0,1/2)$ is the Hurst parameter, σ is a continuous function and $W = \rho B + \sqrt{1-\rho^2}B^{\perp}$ with $(B, B^{\perp})$ a two-dimensional $\mathbb{F}$-Brownian motion and $\rho \in [-1,1]$. Moreover, θ is a deterministic mean reversion level allowed to be time-dependent to fit the market forward variance curve $(\mathbb{E}[V_t])_{t\le T}$ as explained in Section 2 and in Chapter III. Under some general assumptions, we establish in Section 2 the existence of a weak non-negative solution to the fractional stochastic integral equation in (2) exhibiting $H-\varepsilon$ Hölder regularity for any $\varepsilon > 0$. Hence, this class of models is a natural rough extension of classical stochastic volatility models where the fractional kernel is introduced in the drift and stochastic part of the variance process V. Indeed, when $H = 1/2$, we recover classical stochastic volatility models where the variance process is a standard diffusion.

Despite the fit to the historical and implied volatility, some difficulties are encountered in practice for the simulation of rough volatility models and for pricing and hedging derivatives with them. In fact, due to the introduction of the fractional kernel, we lose the Markovian and semimartingale structure. In order to overcome theses difficulties, we approximate these models by simpler ones that we can use in practice.

In Chapters I, II and III, the rough Heston model (which corresponds to the case of $\sigma(x) = \nu\sqrt{x}$) is built as a limit of microscopic Hawkes-based price models. This allowed the understanding of the microstructural foundations of rough volatility and also led to the formula of the characteristic function of the log-price. Hence, the Hawkes approximation enabled us to solve the pricing and hedging under the rough Heston model. However, this approach is specific to the rough Heston case and can not be extended to an arbitrary rough volatility model of the form (1)-(2).

Inspired by the works of [CC98, CCM00, HS15, Mur11] and Chapter IV, we provide a natural Markovian approximation for the class of rough volatility models (1)-(2). The main idea is to write the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ as a Laplace transform of a positive measure μ

$$K(t) = \int_0^\infty e^{-\gamma t}\mu(d\gamma); \quad \mu(d\gamma) = \frac{\gamma^{-H-\frac{1}{2}}}{\Gamma(H+1/2)\Gamma(1/2-H)}d\gamma. \tag{3}$$

We then approximate μ by a finite sum of Dirac measures $\mu^n = \sum_{i=1}^n c_i^n \delta_{\gamma_i^n}$ with positive weights $(c_i^n)_{1\le i\le n}$ and mean reversions $(\gamma_i^n)_{1\le i\le n}$, for $n \ge 1$. This in turn yields an approximation of the fractional kernel by a sequence of smoothed kernels $(K^n)_{n\geq 1}$ given by

$$K^n(t) = \sum_{i=1}^{n} c_i^n e^{-\gamma_i^n t}, \quad n \geq 1.$$

This leads to a multi-factor stochastic volatility model $(S^n, V^n) = (S_t^n, V_t^n)_{t\leq T}$, which is Markovian with respect to the spot price and n variance factors $(V^{n,i})_{1\leq i\leq n}$ and is defined as follows

$$dS_t^n = S_t^n \sqrt{V_t^n} dW_t, \quad V_t^n = g^n(t) + \sum_{i=1}^{n} c_i^n V_t^{n,i}, \tag{4}$$

where

$$dV_t^{n,i} = (-\gamma_i^n V_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n) dB_t,$$

and $g^n(t) = V_0 + \int_0^t K^n(t-s)\theta(s)ds$ with the initial conditions $S_0^n = S_0$ and $V_0^{n,i} = 0$. Note that the factors $(V^{n,i})_{1\leq i\leq n}$ share the same dynamics except that they mean revert at different speeds $(\gamma_i^n)_{1\leq i\leq n}$. Relying on existence results of stochastic Volterra equations in [AJLP17] and Chapter IV, we provide in Theorem 1 the strong existence and uniqueness of the model (S^n, V^n), under some general conditions. Thus the approximation (4) is uniquely well-defined. We can therefore deal with simulation, pricing and hedging problems under these multi-factor models by using standard methods developed for stochastic volatility models.

Theorem 2, which is the main result of this paper, establishes the convergence of the multi-factor approximation sequence $(S^n, V^n)_{n\geq 1}$ to the rough volatility model (S, V) in (1)-(2) when the number of factors n goes to infinity, under a suitable choice of the weights and mean reversions $(c_i^n, \gamma_i^n)_{1\leq i\leq n}$. This convergence is obtained from a general result about stability of stochastic Volterra equations derived in Section 3.4.

In [AJLP17] and Chapters II and III, the characteristic function of the log-price for the specific case of the rough Heston model is obtained in terms of a solution of a fractional Riccati equation. We highlight in Section 4.1 that the corresponding multi-factor approximation (4) inherits a similar affine structure as in the rough Heston model. More precisely, it displays the same characteristic function formula involving a n-dimensional classical Riccati equation instead of the fractional one. This suggests solving numerically the fractional Riccati equation by approximating it through a n-dimensional classical Riccati equation with large n, see Theorem 4. In Section 4.2, we discuss the accuracy and complexity of this numerical method and compare it to the Adams scheme, see [DFF02, DFF04, DF99] and Chapter II.

The paper is organized as follows. In Section 2, we define the class of rough volatility models (1)-(2) and discuss the existence of such models. Then, in Section 3, we build a sequence of multi-factor stochastic volatility models of the form of (4) and show its convergence to a rough volatility model. By applying this approximation to the specific case of the rough Heston model, we obtain a numerical method for computing solutions of fractional Riccati equations that is discussed in Section 4. Finally, some proofs are relegated to Section 5 and some useful technical results are given in an Appendix.

2 A definition of rough volatility models

We provide in this section the precise definition of rough volatility models given by (1)-(2). We discuss the existence of such models and more precisely of a non-negative solution of the fractional stochastic integral equation (2). The existence of an unconstrained weak solution $V = (V_t)_{t\le T}$ is guaranteed by Corollary 2 in the Appendix when σ is a continuous function with linear growth and θ satisfies the condition

$$\forall \varepsilon > 0, \quad \exists C_\varepsilon > 0; \quad \forall u \in (0,T] \quad |\theta(u)| \le C_\varepsilon u^{-\frac{1}{2}-\varepsilon}. \tag{5}$$

Furthermore, the paths of V are Hölder continuous of any order strictly less than H and

$$\sup_{t\in[0,T]} \mathbb{E}[|V_t|^p] < \infty, \quad p > 0. \tag{6}$$

Moreover using Theorem 6 together with Remarks 4 and 5 in the Appendix[1], the existence of a non-negative continuous process V satisfying (2) is obtained under the additional conditions of non-negativity of V_0 and θ and $\sigma(0) = 0$. We can therefore introduce the following class of rough volatility models.

Definition 1. *(Rough volatility models) We define a rough volatility model by any $\mathbb{R}\times\mathbb{R}_+$-valued continuous process $(S,V) = (S_t, V_t)_{t\le T}$ satisfying*

$$dS_t = S_t\sqrt{V_t}dW_t,$$

$$V_t = V_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}(\theta(u) - \lambda V_u)du + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}\sigma(V_u)dB_u,$$

on a filtred probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ with non-negative initial conditions (S_0, V_0). Here T is a positive time horizon, the parameter λ is non-negative, $H \in (0,1/2)$ is the Hurst parameter and $W = \rho B + \sqrt{1-\rho^2}B^\perp$ with $(B, B^\perp)$ a two-dimensional $\mathbb{F}$-Brownian motion and $\rho \in [-1,1]$. Moreover, to guarantee the existence of such model, $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is assumed continuous with linear growth such that $\sigma(0) = 0$ and $\theta : [0,T] \mapsto \mathbb{R}$ is a deterministic non-negative function satisfying (5).

As done in Chapter III, we allow the mean reversion level θ to be time dependent in order to be consistent with the market forward variance curve. More precisely, the following result shows that the mean reversion level θ can be written as a functional of the forward variance curve $(\mathbb{E}[V_t])_{t\le T}$.

Proposition 1. *Let (S,V) be a rough volatility model given by Definition 1. Then, $(\mathbb{E}[V_t])_{t\le T}$ is linked to θ by the following formula*

$$\mathbb{E}[V_t] = V_0 + \int_0^t (t-s)^{\alpha-1}E_\alpha(-\lambda(t-s)^\alpha)\theta(s)ds, \quad t \in [0,T], \tag{7}$$

[1] Theorem 6 is used here with the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ together with $b(x) = -\lambda x$ and $g(t) = V_0 + \int_0^t K(t-u)\theta(u)du$.

where $\alpha = H + 1/2$ *and* $E_\alpha(x) = \sum_{k\geq 0} \frac{x^k}{\Gamma(\alpha(k+1))}$ *is the Mittag-Leffler function. Moreover,* $(\mathbb{E}[V_t])_{t\leq T}$ *admits a fractional derivative*[2] *of order* α *at each time* $t \in (0,T]$ *and*

$$\theta(t) = D^\alpha(\mathbb{E}[V_\cdot] - V_0)_t + \lambda\mathbb{E}[V_t], \quad t \in (0,T]. \tag{8}$$

Proof. Thanks to (6) together with Fubini theorem, $t \mapsto \mathbb{E}[V_t]$ solves the following fractional linear integral equation

$$\mathbb{E}[V_t] = V_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-\frac{1}{2}}(\theta(s) - \lambda\mathbb{E}[V_s])ds, \quad t \in [0,T], \tag{9}$$

yielding (7) by Theorem 5 and Remark 3 in the Appendix. Finally, (8) is obviously obtained from (9). □

Finally, note that uniqueness of the fractional stochastic integral equation (2) is a difficult problem. Adapting the proof in [MS15], we can prove pathwise uniqueness when σ is η-Hölder continuous with $\eta \in (1/(1+2H), 1]$. This result does not cover the square-root case, i.e. $\sigma(x) = \nu\sqrt{x}$, for which weak uniqueness has been established in [AJLP17, MS15], see also Chapter IV.

3 Multi-factor approximation of rough volatility models

Thanks to the small Hölder regularity of the variance process, models of Definition 1 are able to reproduce the rough behavior of the volatility observed in a wide range of assets. However, the fractional kernel forces the variance process to leave both the semimartingale and Markovian worlds, which makes numerical approximation procedures a difficult and challenging task in practice. The aim of this section is to construct a tractable and satisfactory Markovian approximation of any rough volatility model (S,V) of Definition 1. Because S is entirely determined by $(\int_0^\cdot V_s ds, \int_0^\cdot \sqrt{V_s}dW_s)$, it suffices to construct a suitable approximation of the variance process V. This is done by smoothing the fractional kernel.

More precisely, denoting by $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$, the fractional stochastic integral equation (2) reads

$$V_t = V_0 + \int_0^t K(t-s)\left((\theta(s) - \lambda V_s)ds + \sigma(V_s)dB_s\right),$$

which is a stochastic Volterra equation. Approximating the fractional kernel K by a sequence of smooth kernels $(K^n)_{n\geq 1}$, one would expect the convergence of the following corresponding sequence of stochastic Volterra equations

$$V_t^n = V_0 + \int_0^t K^n(t-s)\left((\theta(s) - \lambda V_s^n)ds + \sigma(V_s^n)dB_s\right), \quad n \geq 1,$$

to the fractional one.

[2] Recall that the fractional derivative of order $\alpha \in (0,1)$ of a function f is given by $\frac{d}{dt}\int_0^t \frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)} f(s)ds$ whenever this expression is well defined.

The argument of this section runs as follows. First, exploiting the identity (3), we construct a family of potential candidates for $(K^n, V^n)_{n\geq 1}$ in Section 3.1 such that V^n enjoys a Markovian structure. Second, we provide convergence conditions of $(K^n)_{n\geq 1}$ to K in $\mathbb{L}^2([0,T],\mathbb{R})$ in Section 3.2. Finally, the approximation result for the rough volatility model (S,V) is established in Section 3.3 relying on an abstract stability result of stochastic Volterra equations postponed to Section 3.4 for sake of exposition.

3.1 Construction of the approximation

In [CC98, HS15, Mur11], a Markovian representation of the fractional Brownian motion of Riemann-Liouville type is provided by writing the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ as a Laplace transform of a non-negative measure μ as in (3). This representation is extended in Chapter IV for the Volterra square-root process. Adopting the same approach, we establish a similar representation for any solution of the fractional stochastic integral equation (2) in terms of an infinite dimensional system of processes sharing the same Brownian motion and mean reverting at different speeds. Indeed by using the linear growth of σ together with the stochastic Fubini theorem, see [Ver12], we obtain that

$$V_t = g(t) + \int_0^\infty V_t^\gamma \mu(d\gamma), \quad t \in [0,T],$$

with

$$dV_t^\gamma = (-\gamma V_t^\gamma - \lambda V_t)dt + \sigma(V_t)dB_t, \quad V_0^\gamma = 0, \quad \gamma \geq 0,$$

and

$$g(t) = V_0 + \int_0^t K(t-s)\theta(s)ds. \tag{10}$$

Inspired by [CC98, CCM00], we approximate the measure μ by a weighted sum of Dirac measures

$$\mu^n = \sum_{i=1}^n c_i^n \delta_{\gamma_i^n}, \quad n \geq 1,$$

leading to the following approximation $V^n = (V_t^n)_{t\leq T}$ of the variance process V

$$V_t^n = g^n(t) + \sum_{i=1}^n c_i^n V_t^{n,i}, \quad t \in [0,T], \tag{11}$$

$$dV_t^{n,i} = (-\gamma_i^n V_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n)dB_t, \quad V_0^{n,i} = 0,$$

where

$$g^n(t) = V_0 + \int_0^t K^n(t-u)\theta(u)du, \tag{12}$$

and

$$K^n(t) = \sum_{i=1}^n c_i^n e^{-\gamma_i^n t}. \tag{13}$$

The choice of the positive weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(\gamma_i^n)_{1\leq i\leq n}$, which is crucial for the accuracy of the approximation, is studied in Section 3.2 below. Before proving the convergence of $(V^n)_{n\geq 1}$, we shall first discuss the existence and uniqueness of such processes. This is done by rewriting the stochastic equation (11) as a stochastic Volterra equation of the form

$$V_t^n = g^n(t) + \int_0^t K^n(t-s)\big(-\lambda V_s^n ds + \sigma(V_s^n) dB_s\big), \quad t \in [0,T]. \tag{14}$$

The existence of a continuous non-negative weak solution V^n is ensured by Theorem 6 together with Remarks 4 and 5 in the Appendix[3], because θ and V_0 are non-negative and $\sigma(0) = 0$. Moreover, pathwise uniqueness of solutions to (14) follows by adapting the standard arugments of [YW71], provided a suitable Hölder continuity of σ, see Proposition 8 in the Appendix. Note that this extension is made possible due to the smoothness of the kernel K^n. For instance, this approach fails for the fractional kernel because of the singularity at zero. This leads us to the following result which establishes the strong existence and uniqueness of a non-negative solution of (14) and equivalently of (11).

Theorem 1. *Assume that $\theta : [0,T] \mapsto \mathbb{R}$ is a deterministic non-negative function satisfying* (5) *and that $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is η-Hölder continuous with $\sigma(0) = 0$ and $\eta \in [1/2, 1]$. Then, there exists a unique strong non-negative solution $V^n = (V_t^n)_{t\leq T}$ to the stochastic Volterra equation* (14) *for each $n \geq 1$.*

Due to the uniqueness of (11), we obtain that V^n is a Markovian process according to n state variables $(V^{n,i})_{1\leq i\leq n}$ that we call the factors of V^n. Moreover, V^n being non-negative, it can model a variance process. This leads to the following definition of multi-factor stochastic volatility models.

Definition 2. *(Multi-factor stochastic volatility models). We define the following sequence of multi-factor stochastic volatility models $(S^n, V^n) = (S_t^n, V_t^n)_{t\leq T}$ as the unique $\mathbb{R}\times\mathbb{R}_+$-valued strong solution of*

$$dS_t^n = S_t^n \sqrt{V_t^n} dW_t, \quad V_t^n = g^n(t) + \sum_{i=1}^n c_i^n V_t^{n,i},$$

with

$$dV_t^{n,i} = (-\gamma_i^n V_t^{n,i} - \lambda V_t^n) dt + \sigma(V_t^n) dB_t, \quad V_0^{n,i} = 0, \quad S_0^n = S_0 > 0,$$

on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$, where $\mathbb{F}$ is the canonical filtration a two-dimensional Brownian motion $(W, W^\perp)$ and $B = \rho W + \sqrt{1-\rho^2} W^\perp$ with $\rho \in [-1,1]$. Here, the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(\gamma_i^n)_{1\leq i\leq n}$ are positive, $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is η-Hölder continuous such that $\sigma(0) = 0$, $\eta \in [1/2, 1]$ and g^n is given by (12), *that is*

$$g^n(t) = V_0 + \int_0^t K^n(t-s)\theta(s) ds,$$

with a non-negative initial variance V_0, a kernel K^n defined as in (13) *and a non-negative deterministic function $\theta : [0,T] \mapsto \mathbb{R}$ satisfying* (5).

Note that the strong existence and uniqueness of (S^n, V^n) follows from Theorem 1. This model is Markovian with $n+1$ state variables which are the spot price S^n and the factors of the variance process $V^{n,i}$ for $i \in \{1, \ldots, n\}$.

[3] Theorem 6 is used here with the smoothed kernel K^n given by (13) together with $b(x) = -\lambda x$ and g defined as in (10)

3.2 An approximation of the fractional kernel

Relying on (14), we can see the process V^n as an approximation of V, solution of (2), obtained by smoothing the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ into $K^n(t) = \sum_{i=1}^n c_i^n e^{-\gamma_i^n t}$. Intuitively, we need to choose K^n close to K when n goes to infinity, so that $(V^n)_{n\geq 1}$ converges to V. Inspired by [CCM00], we give in this section a condition on the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversion terms $0 < \gamma_1^n < \ldots < \gamma_n^n$ so that the following convergence

$$\|K^n - K\|_{2,T} \to 0,$$

holds as n goes to infinity, where $\|\cdot\|_{2,T}$ is the usual $\mathbb{L}^2([0,T],\mathbb{R})$ norm. Let $(\eta_i^n)_{0\leq i\leq n}$ be auxiliary mean reversion terms such that $\eta_0^n = 0$ and $\eta_{i-1}^n \leq \gamma_i^n \leq \eta_i^n$ for $i \in \{1,\ldots,n\}$. Writing K as the Laplace transform of μ as in (3), we obtain that

$$\|K^n - K\|_{2,T} \leq \int_{\eta_n^n}^{\infty} \|e^{-\gamma(\cdot)}\|_{2,T}\mu(d\gamma) + \sum_{i=1}^n J_i^n,$$

with $J_i^n = \|c_i^n e^{-\gamma_i^n(\cdot)} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-\gamma(\cdot)}\mu(d\gamma)\|_{2,T}$. We start by dealing with the first term,

$$\int_{\eta_n^n}^{\infty} \|e^{-\gamma(\cdot)}\|_{2,T}\mu(d\gamma) = \int_{\eta_n^n}^{\infty} \sqrt{\frac{1-e^{-2\gamma T}}{2\gamma}}\mu(d\gamma) \leq \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}}(\eta_n^n)^{-H}.$$

Moreover by choosing

$$c_i^n = \int_{\eta_{i-1}^n}^{\eta_i^n} \mu(d\gamma), \quad \gamma_i^n = \frac{1}{c_i^n}\int_{\eta_{i-1}^n}^{\eta_i^n} \gamma\mu(d\gamma), \quad i \in \{1,\ldots,n\}, \tag{15}$$

and using the Taylor-Lagrange inequality up to the second order, we obtain

$$\left| c_i^n e^{-\gamma_i^n t} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-\gamma t}\mu(d\gamma)\right| \leq \frac{t^2}{2}\int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma - \gamma_i^n)^2\mu(d\gamma), \quad t \in [0,T]. \tag{16}$$

Therefore,

$$\sum_{i=1}^n J_i^n \leq \frac{T^{5/2}}{2\sqrt{5}}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma_i^n - \gamma)^2\mu(d\gamma).$$

This leads to the following inequality

$$\|K^n - K\|_{2,T} \leq f_n^{(2)}\big((\eta_i)_{0\leq i\leq n}\big),$$

where $f_n^{(2)}$ is a function of the auxiliary mean reversions defined by

$$f_n^{(2)}((\eta_i^n)_{1\leq i\leq n}) = \frac{T^{\frac{5}{2}}}{2\sqrt{5}}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma - \gamma_i^n)^2\mu(d\gamma) + \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}}(\eta_n^n)^{-H}. \tag{17}$$

Hence, we obtain the convergence of K^n to the fractional kernel under the following choice of weights and mean reversions.

Assumption 1. *We assume that the weights and mean reversions are given by* (15) *such that* $\eta_0^n = 0 < \eta_1^n < \ldots < \eta_n^n$ *and*

$$\eta_n^n \to \infty, \quad \sum_{i=1}^{n} \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma_i^n - \gamma)^2 \mu(d\gamma) \to 0, \tag{18}$$

as n goes to infinity.

Proposition 2. *Fix $(c_i^n)_{1\le i\le n}$ and $(\gamma_i^n)_{1\le i\le n}$ as in Assumption 1 and K^n given by* (13), *for all $n \ge 1$. Then, $(K^n)_{n\ge1}$ converges in $\mathbb{L}^2[0,T]$ to the fractional kernel $K(t) = \frac{t^{H-1/2}}{\Gamma(H+\frac{1}{2})}$ as n goes to infinity.*

There exists several choices of auxiliary factors such that condition (18) is met. For instance, assume that $\eta_i^n = i\pi_n$ for each $i \in \{0,\ldots,n\}$ such that $\pi_n > 0$. It follows from

$$\sum_{i=1}^{n} \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma - \gamma_i)^2 \mu(d\gamma) \le \pi_n^2 \int_0^{\eta_n^n} \mu(d\gamma) = \frac{1}{(1/2-H)\Gamma(H+1/2)\Gamma(1/2-H)} \pi_n^{\frac{5}{2}-H} n^{\frac{1}{2}-H},$$

that (18) is satisfied for

$$\eta_n^n = n\pi_n \to \infty, \quad \pi_n^{\frac{5}{2}-H} n^{\frac{1}{2}-H} \to 0,$$

as n tends to infinity. In this case,

$$\|K^n - K\|_{2,T} \le \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}} \left((\eta_n^n)^{-H} + \frac{HT^{\frac{5}{2}}}{\sqrt{10}(1/2-H)} \pi_n^2 (\eta_n^n)^{\frac{1}{2}-H} \right).$$

This upper bound is minimal for

$$\pi_n = \frac{n^{-\frac{1}{5}}}{T} \left(\frac{\sqrt{10}(1-2H)}{5-2H} \right)^{\frac{2}{5}}, \tag{19}$$

and

$$\|K^n - K\|_{2,T} \le C_H n^{-\frac{4H}{5}},$$

where C_H is a positive constant that can be computed explicitly and that depends only on the Hurst parameter $H \in (0,1/2)$.

Remark 1. *Note that the kernel approximation in Proposition 2 can be easily extended to any kernel of the form*

$$K(t) = \int_0^\infty e^{-\gamma t} \mu(d\gamma),$$

where μ is a non-negative measure such that

$$\int_0^\infty (1 \wedge \gamma^{-1/2}) \mu(d\gamma) < \infty.$$

3.3 Convergence result

We assume now that the weights and mean reversions of the multi-factor stochastic volatility model (S^n, V^n) satisfy Assumption 1. Thanks to Proposition 2, the smoothed kernel K^n is close to the fractional one for large n. Because V^n satisfies the stochastic Volterra equation (14), V^n has to be close to V and thus by passing to the limit, $(S^n, V^n)_{n\geq 1}$ should converge to the rough volatility model (S, V) of Definition 1 as n goes large. This is the object of the next theorem, which is the main result of this paper.

Theorem 2. *Let $(S^n, V^n)_{n\geq 1}$ be a sequence of multi-factor stochastic volatility models given by Definition 2. Then, under Assumption 1, the family $(S^n, V^n)_{n\geq 1}$ is tight for the uniform topology and any point limit (S, V) is a rough volatility model given by Definition 1.*

Theorem 2 states the convergence in law of $(S^n, V^n)_{n\geq 1}$ whenever the fractional stochastic integral equation (2) admits a unique weak solution. In order to prove Theorem 2, whose proof is in Section 5.2 below, a more general stability result for d-dimensional stochastic Volterra equations is established in the next subsection.

3.4 Stability of stochastic Volterra equations

As mentioned above, Theorem 2 relies on the study of the stability of more general d-dimensional stochastic Volterra equations of the form

$$X_t = g(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \in [0, T], \tag{20}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous and satisfy the linear growth condition, $K \in \mathbb{L}^2([0, T], \mathbb{R}^{d\times d})$ admits a resolvent of the first kind L, see Appendix VIII.A.2, and W is a m-dimensional Brownian motion on some filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$. From Proposition 7 in the Appendix, $g : [0, T] \mapsto \mathbb{R}^d$ and $K \in \mathbb{L}^2([0, T], \mathbb{R}^{d\times d})$ should satisfy Assumption 3, that is

$$|g(t+h) - g(t)|^2 + \int_0^h |K(s)|^2 ds + \int_0^{T-h} |K(h+s) - K(s)|^2 ds \leq C h^{2\gamma}, \tag{21}$$

for any $t, h \geq 0$ with $t + h \leq T$ and for some positive constants C and γ, to guarantee the weak existence of a continuous solution X of (20).

More precisely, we consider a sequence $X^n = (X_t^n)_{t\leq T}$ of continuous weak solutions to the stochastic Volterra equation (20) with a kernel $K^n \in \mathbb{L}^2([0, T], \mathbb{R}^{d\times d})$ admitting a resolvent of the first kind, on some filtered probability space $(\Omega^n, \mathscr{F}^n, \mathbb{F}^n, \mathbb{P}^n)$,

$$X_t^n = g^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s^n, \quad t \in [0, T],$$

with $g^n : [0, T] \mapsto \mathbb{R}^d$ and K^n satisfying (21) for every $n \geq 1$. The stability of (20) means the convergence in law of the family of solutions $(X^n)_{n\geq 1}$ to a limiting process X which is a solution to (20), when (K^n, g^n) is close to (K, g) as n goes large.

This convergence is established by verifying first the Kolmogorov tightness criterion for the sequence $(X^n)_{n\geq 1}$. It is obtained when g^n and K^n satisfy (21) uniformly in n in the following sense.

Assumption 2. *There exists positive constants γ and C such that*

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

for any $t, h \geq 0$ with $t+h \leq T$,

The following result, whose proof is postponed to Section 5.1 below, states the convergence of $(X^n)_{n\geq 1}$ to a solution of (20).

Theorem 3. *Assume that*

$$\int_0^T |K(s)-K^n(s)|^2 ds \longrightarrow 0, \quad g_n(t) \longrightarrow g(t),$$

for any $t \in [0, T]$ as n goes to infinity. Then, under Assumption 2, 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 (20).

4 The particular case of the rough Heston model

The rough Heston model introduced in Chapters II and III is a particular case of the class of rough volatility models of Definition 1, with $\sigma(x) = \nu\sqrt{x}$ for some positive parameter ν, that is

$$dS_t = S_t\sqrt{V_t}dW_t, \quad S_0 > 0,$$
$$V_t = g(t) + \int_0^t K(t-s)\left(-\lambda V_s ds + \nu\sqrt{V_s}dB_s\right),$$

where $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ denotes the fractional kernel and g is given by (10). Aside from reproducing accurately the historical and implied volatility, the rough Heston model displays a closed formula for the characteristic function of the log-price in terms of a solution to a fractional Riccati equation allowing to fast pricing and calibration, see Chapter V. More precisely, it is shown in Chapters II and III that

$$L(t,z) = \mathbb{E}\left[\exp\left(z\log(S_t/S_0)\right)\right]$$

is given by

$$\exp\left(\int_0^t F(z,\psi(t-s,z))g(s)ds\right), \tag{22}$$

where $\psi(\cdot, z)$ is the unique continuous solution of the fractional Riccati equation

$$\psi(t,z) = \int_0^t K(t-s)F(z,\psi(s,z))ds, \quad t \in [0,T], \tag{23}$$

with $F(z,x) = \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2$ and $z \in \mathbb{C}$ such that $\Re(z) \in [0,1]$. Unlike the classical case $H = 1/2$, (23) does not exhibit an explicit solution. However, it can be solved numerically through the Adam scheme developed in [DFF02, DFF04, DF99] and Chapter II for instance. In this section, we show that the multi-factor approximation applied to the rough Heston model gives rise to another natural numerical scheme for solving the fractional Riccati equation. Furthermore, we will establish the convergence of this scheme with explicit errors.

4.1 Multi-factor scheme for the fractional Riccati equation

We consider the multi-factor approximation (S^n, V^n) of Definition 2 with $\sigma(x) = \nu\sqrt{x}$, where the number of factors n is large, that is

$$dS_t^n = S_t^n \sqrt{V_t^n} dW_t, \quad V_t^n = g^n(t) + \sum_{i=1}^n c_i^n V_t^{n,i},$$

with

$$dV_t^{n,i} = (-\gamma_i^n V_t^{n,i} - \lambda V_t^n)dt + \nu\sqrt{V_t^n} dB_t, \quad V_0^{n,i} = 0, \quad S_0^n = S_0.$$

Recall that g^n is given by (12) and it converges pointwise to g as n goes large, see Lemma 1.

We write the dynamics of (S^n, V^n) in terms of a Volterra Heston model with the smoothed kernel K^n given by (13) as follows

$$dS_t^n = S_t^n \sqrt{V_t^n} dW_t,$$

$$V_t^n = g^n(t) - \int_0^t K^n(t-s)\lambda V_s^n ds + \int_0^t K^n(t-s)\nu\sqrt{V_s^n} dB_s.$$

In [AJLP17] and Chapter IV, the characteristic function formula of the log-price (22) is extended to the general class of Volterra Heston models. In particular,

$$L^n(t,z) = \mathbb{E}\left[\exp\left(z\log(S_t^n/S_0)\right)\right]$$

is given by

$$\exp\left(\int_0^t F(z, \psi^n(t-s,z)) g^n(s) ds\right), \tag{24}$$

where $\psi^n(\cdot, z)$ is the unique continuous solution of the Riccati Volterra equation

$$\psi^n(t,z) = \int_0^t K^n(t-s) F(z, \psi^n(s,z)) ds, \quad t \in [0,T], \tag{25}$$

for each $z \in \mathbb{C}$ with $\Re(z) \in [0,1]$.

Thanks to the weak uniqueness of the rough Heston model, established in several works [AJLP17, MS15], and to Theorem 2, $(S^n, V^n)_{n\geq 1}$ converges in law for the uniform topology to (S, V) when n tends to infinity. In particular, $L^n(t,z)$ converges pointwise to $L(t,z)$. Therefore, we expect $\psi^n(\cdot, z)$ to be close to the solution of the fractional Riccati equation (23). This is the object of the next theorem, whose proof is reported to Section 5.3 below.

Theorem 4. *There exists a positive constant C such that, for any $a \in [0,1]$, $b \in \mathbb{R}$ and $n \geq 1$,*

$$\sup_{t \in [0,T]} |\psi^n(t, a+ib) - \psi(t, a+ib)| \leq C(1+b^4) \int_0^T |K^n(s) - K(s)| ds,$$

where $\psi(\cdot, a+ib)$ (resp. $\psi^n(\cdot, a+ib)$) denotes the unique continuous solution of the Riccati Volterra equation (23) *(resp.* (25)*).*

Relying on the L^1-convergence of $(K^n)_{n\geq 1}$ to K under Assumption 1, see Proposition 2, we have the uniform convergence of $(\psi^n(\cdot, z))_{n\geq 1}$ to $\psi(\cdot, z)$ on $[0, T]$. Hence, Theorem 4 suggests a new numerical method for the computation of the fractional Riccati solution (23) where an explicit error is given. Indeed, set

$$\psi^{n,i}(t,z) = \int_0^t e^{-\gamma_i^n (t-s)} F(z, \psi^n(s,z)) ds, \quad i \in \{1, \ldots, n\}.$$

Then,

$$\psi^n(t,z) = \sum_{i=1}^n c_i^n \psi^{n,i}(t,z),$$

and $(\psi^{n,i}(\cdot, z))_{1\leq i\leq n}$ solves the following n-dimensional system of ordinary Riccati equations

$$\partial_t \psi^{n,i}(t,z) = -\gamma_i^n \psi^{n,i}(t,z) + F(z, \psi^n(t,z)), \quad \psi^{n,i}(0,z) = 0, \quad i \in \{1, \ldots, n\}. \tag{26}$$

Hence, (26) can be solved numerically by usual finite difference methods leading to $\psi^n(\cdot, z)$ as an approximation of the fractional Riccati solution.

4.2 Numerical illustrations

In this section, we consider a rough Heston model with the following parameters

$$\lambda = 0.3, \quad \rho = -0.7, \quad \nu = 0.3, \quad H = 0.1, \quad V_0 = 0.02, \quad \theta \equiv 0.02.$$

We discuss the accuracy of the multi-factor approximation sequence $(S^n, V^n)_{n\geq 1}$ as well as the corresponding Riccati Volterra solution $(\psi^n(\cdot, z))_{n\geq 1}$, for different choices of the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(\gamma_i^n)_{1\leq i\leq n}$. This is achieved by first computing, for different number of factors n, the implied volatility $\sigma^n(k, T)$ of maturity T and log-moneyness k by a Fourier inversion of the characteristic function formula (24), see [CM99, Lew01] for instance. In a second step, we compare $\sigma^n(k, T)$ to the implied volatility $\sigma(k, T)$ of the rough Heston model. We also compare the Riccati Volterra solution $\psi^n(T, z)$ to the fractional one $\psi(T, z)$.

Note that the Riccati Volterra solution $\psi^n(\cdot, z)$ is computed by solving numerically the n-dimensional Riccati equation (26) with a classical finite difference scheme. The complexity of such scheme is $O(n \times n_{\Delta t})$, where $n_{\Delta t}$ is the number of time steps applied for the scheme, while the complexity of the Adam scheme used for the computation of $\psi(\cdot, z)$ is $O(n_{\Delta t}^2)$. In the following numerical illustrations, we fix $n_{\Delta t} = 200$.

In order to guarantee the convergence, the weights and mean reversions have to satisfy Assumption 1 and in particular they should be of the form (15) in terms of auxiliary mean reversions $(\eta_i^n)_{0\leq i\leq n}$ satisfying (18). For instance, one can fix

$$\eta_i^n = i\pi_n, \quad i \in \{0,\ldots,n\}, \tag{27}$$

where π_n is defined by (19), as previously done in Section 3.2. For this particular choice, Figure VIII.1 shows a decrease of the relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ towards zero for different values of b.

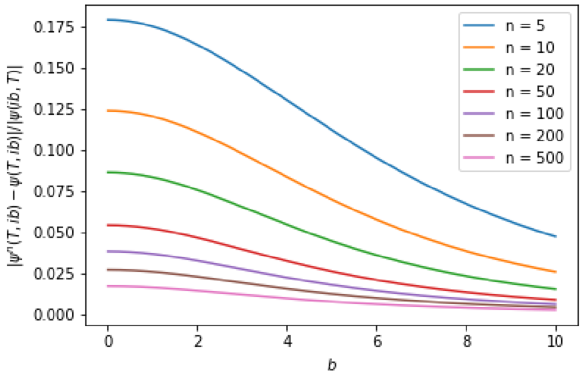

Figure VIII.1 – The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (27) and for different numbers of factors n with $T=1$.

We also observe in the Figure VIII.2 below that the implied volatility $\sigma^n(k,T)$ of the multi-factor approximation is close to $\sigma(k,T)$ for a number of factors $n\geq 20$. Notice that the approximation is more accurate around the money.

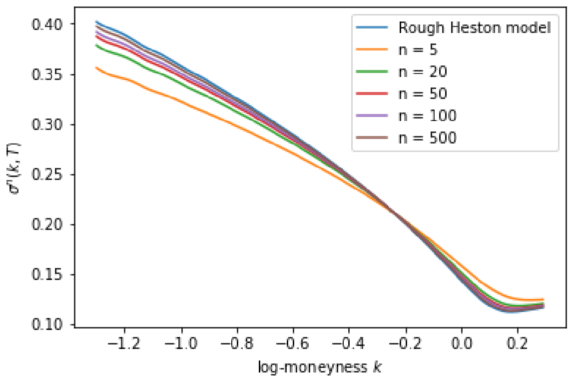

Figure VIII.2 – Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (27) and for different numbers of factors n with $T=1$.

In order to obtain a more accurate convergence, we can minimize the upper bound $f_n^{(2)}((\eta_i^n)_{0\leq i\leq n})$ of $\|K^n-K\|_{2,T}$ defined in (17). Hence, we choose $(\eta_i^n)_{0\leq i\leq n}$ to be a solution of the constrained minimization problem

$$\inf_{(\eta_i^n)_i \in \mathscr{E}_n} f_n^{(2)}((\eta_i^n)_{0\le i\le n}), \tag{28}$$

where $\mathscr{E}_n = \{(\eta_i^n)_{0\le i\le n};\quad 0 = \eta_0^n < \eta_1^n < \ldots < \eta_n^n\}$.

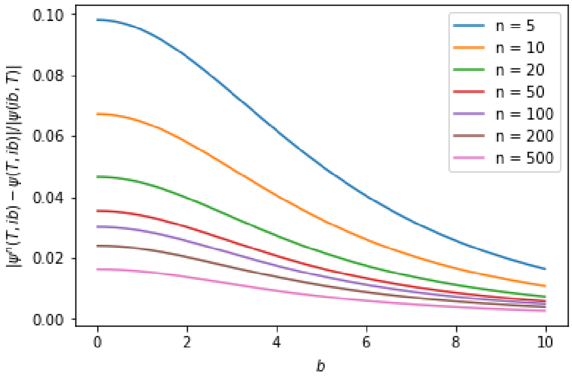

Figure VIII.3 – The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (28) and for different numbers of factors n with $T = 1$.

We notice from Figure VIII.3, that the relative error $|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}|$ is smaller under the choice of factors (28). Indeed the Volterra approximation $\psi^n(T,ib)$ is now closer to the fractional Riccati solution $\psi(T,ib)$ especially for small number of factors. However, when n is large, the accuracy of the approximation seems to be close to the one under (27). For instance when $n = 500$, the relative error is around 1% under both (27) and (28).

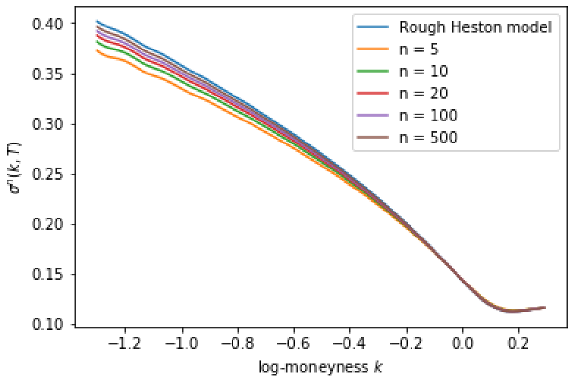

Figure VIII.4 – Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (28) and for different numbers of factors n with $T = 1$.

In the same way, we observe in Figure VIII.4 that the accuracy of the implied volatility approximation $\sigma^n(k,T)$ is more satisfactory under (28) especially for a small number of factors.

Theorem 4 states that the convergence of $\psi^n(\cdot,z)$ depends actually on the $\mathbb{L}^1([0,T],\mathbb{R})$-error between K^n and K. Similarly to the computations of Section 3.2, we may show that,

$$\int_0^T |K^n(s)-K(s)|ds \le f_n^{(1)}((\eta_i^n)_{0\le i\le n}),$$

where

$$f_n^{(1)}((\eta_i^n)_{0\le i\le n}) = \frac{T^3}{6}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma-\gamma_i^n)^2\mu(d\gamma) + \frac{1}{\Gamma(H+3/2)\Gamma(1/2-H)}(\eta_n^n)^{-H-\frac{1}{2}}.$$

This leads to choosing $(\eta_i^n)_{0\le i\le n}$ as a solution of the constrained minimization problem

$$\inf_{(\eta_i^n)_i\in\mathscr{E}_n} f_n^{(1)}((\eta_i^n)_{0\le i\le n}). \tag{29}$$

It is easy to show that such auxiliary mean-reversions $(\eta_i^n)_{0\le i\le n}$ satisfy (18) and thus Assumption 1 is met.

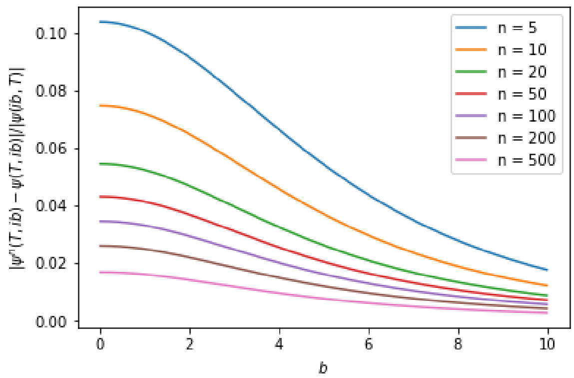

Figure VIII.5 – The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (29) and for different numbers of factors n with $T=1$.

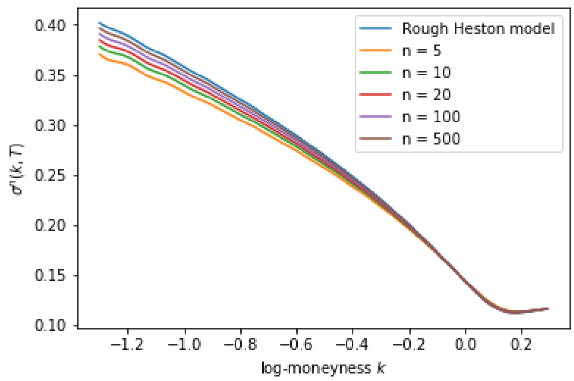

Figure VIII.6 – Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (29) and for different numbers of factors n with $T=1$.

Figures VIII.5 and VIII.6 exhibit similar results as the ones in Figures VIII.3 and VIII.4 corresponding to the choice of factors (28). In fact, we notice in practice that the solution of the minimization problem (28) is close to the one in (29).

4.3 Upper bound for call prices error

Using a Fourier transform method, we can also provide an error between the price of the call $C^n(k,T) = \mathbb{E}[(S_T^n - S_0 e^k)_+]$ in the multi-factor model and the price of the same call $C(k,T) = \mathbb{E}[(S_T - S_0 e^k)_+]$ in the rough Heston model. However, for technical reasons, this bound is obtained for a modification of the multi-factor approximation $(S^n, V^n)_{n\geq 1}$ of Definition 2 where the function g^n given initially by (12) is updated into

$$g^n(t) = \int_0^t K^n(t-s)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2-H)} + \theta(s)\big) ds, \tag{30}$$

where K^n is the smoothed approximation (13) of the fractional kernel. Note that the strong existence and uniqueness of V^n is still directly obtained from Proposition 8 and its non-negativity from Theorem 6 together with Remarks 4 and 5 in the Appendix[4]. Although for g^n satisfying (30), $(V^n)_{n\geq 1}$ can not be tight[5], the corresponding spot price $(S^n)_{n\geq 1}$ converges as shown in the following proposition.

Proposition 3. *Let $(S^n, V^n)_{n\geq 1}$ be a sequence of multi-factor Heston models as in Definition 2 with $\sigma(x) = \nu\sqrt{x}$ and g^n given by* (30). *Then, under Assumption 1, $(S^n, \int_0^\cdot V_s^n ds)_{n\geq 1}$ converges in law for the uniform topology to $(S, \int_0^\cdot V_s ds)$, where (S,V) is a rough Heston model as in Definition 1 with $\sigma(x) = \nu\sqrt{x}$.*

Note that the characteristic function (24) still holds. Using Theorem 4 together with a Fourier transform method, we obtain an explicit error for the call prices. We refer to Section 5.5 below for the proof.

Proposition 4. *Let $C(k,T)$ be the price of the call in the rough Heston model with maturity $T>0$ and log-moneyness $k \in \mathbb{R}$. We denote by $C^n(k,T)$ the price of the call in the multi-factor Heston model of Definition 2 such that g^n is given by* (30). *If $|\rho|<1$, then there exists a positive constant $c>0$ such that*

$$|C(k,T) - C^n(k,T)| \leq c \int_0^T |K(s) - K^n(s)| ds, \qquad n \geq 1.$$

5 Proofs

In this section, we use the convolution notations together with the resolvent definitions of Appendix VIII.A. We denote by c any positive constant independent of the variables t, h and n and that may vary from line to line. For any $h \in \mathbb{R}$, we will use the notation Δ_h to denote

[4] Note that Theorem 6 is used here for the smoothed kernel K^n, $b(x) = -\lambda x$ and g^n defined by (30).

[5] In fact, $V_0^n = 0$ while V_0 may be positive.

the semigroup operator of right shifts defined by $\Delta_h f : t \mapsto f(h+t)$ for any function f.

We first prove Theorem 3, which is the building block of Theorem 2. Then, we turn to the proofs of the results contained in Section 4, which concern the particular case of the rough Heston model.

5.1 Proof of Theorem 3

Tightness of $(X^n)_{n\geq 1}$ **:** We first show that, for any $p \geq 2$,

$$\sup_{n\geq 1} \sup_{t\leq T} \mathbb{E}[|X_t^n|^p] < \infty. \tag{31}$$

Thanks to Proposition 7, we already have

$$\sup_{t\leq T} \mathbb{E}[|X_t^n|^p] < \infty. \tag{32}$$

Using the linear growth of (b,σ) and (32) together with Jensen and BDG inequalities, we get

$$\mathbb{E}[|X_t^n|^p] \leq c\left(\sup_{t\leq T} |g^n(t)|^p + \left(\int_0^T |K^n(s)|^2 ds\right)^{\frac{p}{2}-1} \int_0^t |K^n(t-s)|^2 (1+\mathbb{E}[|X_s^n|^p]) ds\right).$$

Relying on Assumption 2 and the convergence of $(g^n(0), \int_0^T |K^n(s)|^2 ds)_{n\geq 1}$, $\sup_{t\leq T} |g^n(t)|^p$ and $\int_0^T |K^n(s)|^2 ds$ are uniformly bounded in n. This leads to

$$\mathbb{E}[|X_t^n|^p] \leq c\left(1 + \int_0^t |K^n(t-s)|^2 \mathbb{E}[|X_s^n|^p] ds\right).$$

By the Grönwall type inequality in Lemma 5 in the Appendix, we deduce that

$$\mathbb{E}[|X_t^n|^p] \leq c\left(1 + \int_0^t E_c^n(s) ds\right) \leq c\left(1 + \int_0^T E_c^n(s) ds\right),$$

where $E_c^n \in \mathbb{L}^1([0,T],\mathbb{R})$ is the canonical resolvent of $|K^n|^2$ with parameter c, defined in Appendix VIII.A.3, and the last inequality follows from the fact that $\int_0^\cdot E_c^n(s) ds$ is non-decreasing by Corollary 3. The convergence of $|K^n|^2$ to $|K|^2$ in $\mathbb{L}^1([0,T],\mathbb{R})$ implies the convergence of E_c^n to the canonical resolvent of $|K|^2$ with parameter c in $\mathbb{L}^1([0,T],\mathbb{R})$, see [GLS90, Theorem 2.3.1]. Thus, $\int_0^T E_c^n(s) ds$ is uniformly bounded in n, yielding (31).

We now show that $(X^n)_{n\geq 1}$ exhibits the Kolmogorov tightness criterion. In fact, using again the linear growth of (b, σ) and (31) together with Jensen and BDG inequalities, we obtain, for any $p \geq 2$ and $t, h \geq 0$ such that $t+h \leq T$,

$$\mathbb{E}[|X_{t+h}^n - X_t^n|^p] \leq c\Big(|g^n(t+h) - g^n(t)|^p + \Big(\int_0^{T-h} |K^n(h+s) - K^n(s)|^2 ds\Big)^{p/2} + \Big(\int_0^h |K^n(s)|^2 ds\Big)^{p/2}\Big).$$

Hence, Assumption 2 leads to

$$\mathbb{E}[|X_{t+h}^n - X_t^n|^p] \leq c h^{p\gamma},$$

and therefore to the tightness of $(X^n)_{n\geq 1}$ for the uniform topology.

Convergence of $(X^n)_{n\geq 1}$: Let $M_t^n = \int_0^t \sigma(X_s^n)dW_s^n$. As $\langle M^n\rangle_t = \int_0^t \sigma\sigma^*(X_s^n)ds$, $(\langle M^n\rangle)_{n\geq 1}$ is tight and consequently we get the tightness of $(M^n)_{n\geq 1}$ from [JS13, Theorem VI-4.13]. Let $(X,M) = (X_t, M_t)_{t\leq T}$ be a possible limit point of $(X^n, M^n)_{n\geq 1}$. Thanks to [JS13, Theorem VI-6.26], M is a local martingale and necessarily

$$\langle M\rangle_t = \int_0^t \sigma\sigma^*(X_s)ds, \quad t\in[0,T].$$

Moreover, setting $Y_t^n = \int_0^t b(X_s^n)ds + M_t^n$, the assoicativity property (50) in the Appendix yields

$$(L * X^n)_t = (L * g^n)(t) + \left(L * \left((K^n - K) * dY^n\right)\right)_t + Y_t^n, \tag{33}$$

where L is the resolvent of the first kind of K defined in Appendix VIII.A.2. By the Skorokhod representation theorem, we construct a probability space supporting a sequence of copies of $(X^n, M^n)_{n\geq 1}$ that converges uniformly on $[0,T]$, along a subsequence, to a copy of (X,M) almost surely, as n goes to infinity. We maintain the same notations for these copies. Hence, we have

$$\sup_{t\in[0,T]} |X_t^n - X_t| \to 0, \quad \sup_{t\in[0,T]} |M_t^n - M_t| \to 0,$$

almost surely, as n goes to infinity. Relying on the continuity and linear growth of b together with the dominated convergence theorem, it is easy to obtain for any $t\in[0,T]$

$$(L * X^n)_t \to (L * X)_t, \quad \int_0^t b(X_s^n)ds \to \int_0^t b(X_s)ds,$$

almost surely as n goes to infinity. Moreover for each $t\in[0,T]$

$$(L * g^n)(t) \to (L * g)(t),$$

by the uniform boundedness of g^n in n and t and the dominated convergence theorem. Finally thanks to the Jensen inequality,

$$\mathbb{E}[|\left(L * ((K^n - K) * dY^n)\right)_t|^2] \leq c \sup_{t\leq T} \mathbb{E}[|\left((K^n - K) * dY^n\right)_t|^2].$$

From (31) and the linear growth of (b,σ), we deduce

$$\sup_{t\leq T} \mathbb{E}[|\left((K^n - K) * dY^n\right)_t|^2] \leq c\int_0^T |K^n(s) - K(s)|^2 ds,$$

which goes to zero when n is large. Consequently, we send n to infinity in (33) and obtain the following almost surely equality, for each $t\in[0,T]$,

$$(L * X)_t = (L * g)(t) + \int_0^t b(X_s)ds + M_t. \tag{34}$$

Recall also that $\langle M\rangle = \int_0^\cdot \sigma\sigma^*(X_s)ds$. Hence, by [RY13, Theorem V-3.9], there exists a m-dimensional Brownian motion W such that

$$M_t = \int_0^t \sigma(X_s)dW_s, \quad t\in[0,T].$$

The processes in (34) being continuous, we deduce that, almost surely,

$$(L * X)_t = (L * g)(t) + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \quad t \in [0,T].$$

We convolve by K and use the associativity property (50) in the Appendix to get that, almost surely,

$$\int_0^t X_s ds = \int_0^t g(s)ds + \int_0^t \left(\int_0^s K(s-u)(b(X_u)du + \sigma(X_u)dW_u) \right) ds, \quad t \in [0,T].$$

Finally it is easy to see that the processes above are differentiable and we conclude that X is solution of the stochastic Volterra equation (20), by taking the derivative.

5.2 Proof of Theorem 2

Theorem 2 is easily obtained once we prove the tightness of $(V^n)_{n\geq 1}$ for the uniform topology and that any limit point V is solution of the fractional stochastic integral equation (2). This is a direct consequence of Theorem 3, by setting $d = m = 1$, g and g^n respectively as in (10) and (12), $b(x) = -\lambda x$, K being the fractional kernel and $K^n(t) = \sum_{i=1}^n c_i^n e^{-\gamma_i^n t}$ its smoothed approximation. Under Assumption 1, $(K^n)_{n\geq 1}$ converges in $\mathbb{L}^2([0,T],\mathbb{R})$ to the fractional kernel, see Proposition 2. Hence, it is left to show the pointwise convergence of $(g^n)_{n\geq 1}$ to g on $[0,T]$ and that $(K^n, g^n)_{n\geq 1}$ satisfies Assumption 2.

Lemma 1 (Convergence of g^n). *Define $g^n : [0,T] \mapsto \mathbb{R}$ and $g : [0,T] \mapsto \mathbb{R}$ respectively by* (10) *and* (12) *such that $\theta : [0,T] \mapsto \mathbb{R}$ satisfies* (5). *Under assumption* (1), *we have for any $t \in [0,T]$*

$$g^n(t) \to g(t),$$

as n tends to infinity.

Proof. Because θ satisfies (5), it is enough to show that for each $t \in [0,T]$

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} |K^n(s) - K(s)| ds \tag{35}$$

converges to zero as n goes large, for some $\varepsilon > 0$ and K^n given by (13). Using the representation of K as the Laplace transform of μ as in (3), we obtain that (35) is bounded by

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} \int_{\eta_n^n}^\infty e^{-\gamma s}\mu(d\gamma)ds + \sum_{i=1}^n \int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} |c_i^n e^{-\gamma_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-\gamma s}\mu(d\gamma)| ds. \tag{36}$$

The first term in (36) converges to zero for large n by the dominated convergence theorem because η_n^n tends to infinity, see Assumption 1. Using the Taylor-Lagrange inequality (16), the second term in (36) is dominated by

$$\frac{1}{2} \int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} s^2 ds \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (\gamma - \gamma_i^n)^2 \mu(d\gamma),$$

which goes to zero thanks to Assumption 1. □

Lemma 2 (K^n satisfying Assumption 2). *Under Assumption 1, there exists $C>0$ such that, for any $t,h\geq 0$ with $t+h\leq T$,*

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

where K^n is defined by (13).

Proof. We start by proving that for any $t,h\geq 0$ with $t+h\leq T$

$$\int_0^h|K^n(s)|^2ds\leq ch^{2H}. \tag{37}$$

In fact we know that this inequality is satisfied for $K(t)=\frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$. Thus it is enough to prove

$$\|K^n-K\|_{2,h}\leq ch^H,$$

where $\|\cdot\|_{2,h}$ stands for the usual $\mathbb{L}^2([0,h],\mathbb{R})$ norm. Relying on the Laplace transform representation of K given by (3), we obtain

$$\|K^n-K\|_{2,h}\leq\int_{\eta_n^n}^{\infty}\|e^{-\gamma(\cdot)}\|_{2,h}\mu(d\gamma)+\sum_{i=1}^n J_{i,h}^n,$$

where $J_{i,h}^n=\|c_i^n e^{-\gamma_i^n(\cdot)}-\int_{\eta_{i-1}^n}^{\eta_i^n}e^{-\gamma(\cdot)}\mu(d\gamma)\|_{2,h}$. We start by bounding the first term,

$$\begin{aligned}\int_{\eta_n^n}^{\infty}\|e^{-\gamma(\cdot)}\|_{2,h}\mu(d\gamma)&\leq\int_0^{\infty}\sqrt{\frac{1-e^{-2\gamma h}}{2\gamma}}\mu(d\gamma)\\&=\frac{h^H}{\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}}\int_0^{\infty}\sqrt{\frac{1-e^{-2\gamma}}{\gamma}}\gamma^{-H-\frac{1}{2}}d\gamma.\end{aligned}$$

As in Section 3.2, we use the Taylor-Lagrange inequality (16) to get

$$\sum_{i=1}^n J_{i,h}^n\leq\frac{1}{2\sqrt{5}}h^{\frac{5}{2}}\sum_{i=1}^n\int_{\eta_{i-1}^n}^{\eta_i^n}(\gamma-\gamma_i^n)^2\mu(d\gamma).$$

Using the boundedness of $\left(\sum_{i=1}^n\int_{\eta_{i-1}^n}^{\eta_i^n}(\gamma-\gamma_i^n)^2\mu(d\gamma)\right)_{n\geq 1}$ from Assumption 1, we deduce (37). We now prove

$$\int_0^{T-h}|K^n(h+s)-K^n(s)|^2ds\leq ch^{2H}. \tag{38}$$

In the same way, it is enough to show

$$\|(\Delta_h K^n-\Delta_h K)-(K^n-K)\|_{2,T-h}\leq ch^H,$$

Similarly to the previous computations, we get

$$\|(\Delta_h K^n-\Delta_h K)-(K^n-K)\|_{2,T-h}\leq\int_{\eta_n^n}^{\infty}\|e^{-\gamma(\cdot)}-e^{-\gamma(h+\cdot)}\|_{2,T-h}\mu(d\gamma)+\sum_{i=1}^n\widetilde{J}_{i,h}^n,$$

with $\widetilde{J}^n_{i,h} = \|c^n_i(e^{-\gamma^n_i(\cdot)} - e^{-\gamma^n_i(h+\cdot)}) - \int_{\eta^n_{i-1}}^{\eta^n_i}(e^{-\gamma(\cdot)} - e^{-\gamma(h+\cdot)})\mu(d\gamma)\|_{2,T-h}$. Notice that

$$\int_{\eta^n_n}^{\infty} \|e^{-\gamma(\cdot)} - e^{-\gamma(h+\cdot)}\|_{2,T-h}\mu(d\gamma) = \int_{\eta^n_n}^{\infty}(1-e^{-\gamma h})\sqrt{\frac{1-e^{-2\gamma(T-h)}}{2\gamma}}\mu(d\gamma)$$
$$\leq c\int_0^{\infty}(1-e^{-\gamma h})\gamma^{-H-1}d\gamma \leq ch^H.$$

Moreover, fix $h, t > 0$ and set $\chi(\gamma) = e^{-\gamma t} - e^{-\gamma(t+h)}$. The second derivative reads

$$\chi''(\gamma) = h\big(t^2\gamma e^{-\gamma t}\frac{1-e^{-\gamma h}}{\gamma h} - he^{-\gamma(t+h)} - 2te^{-\gamma(t+h)}\big), \quad \gamma > 0. \tag{39}$$

Because $x \mapsto xe^{-x}$ and $x \mapsto \frac{1-e^{-x}}{x}$ are bounded functions on $(0,\infty)$, there exists $C > 0$ independent of $t, h \in [0, T]$ such that

$$|\chi''(\gamma)| \leq Ch, \quad \gamma > 0.$$

The Taylor-Lagrange formula, up to the second order, leads to

$$|c^n_i(e^{-\gamma^n_i t} - e^{-\gamma^n_i(t+h)}) - \int_{\eta^n_{i-1}}^{\eta^n_i}(e^{-\gamma t} - e^{-\gamma(t+h)})\mu(d\gamma)| \leq \frac{C}{2}h\int_{\eta^n_{i-1}}^{\eta^n_i}(\gamma-\gamma^n_i)^2\mu(d\gamma).$$

Thus

$$\sum_{i=1}^n \widetilde{J}^n_{i,h} \leq \frac{C}{2}h\sum_{i=1}^n\int_{\eta^n_{i-1}}^{\eta^n_i}(\gamma-\gamma^n_i)^2\mu(d\gamma).$$

Finally, (38) follows from the boundedness of $\big(\sum_{i=1}^n\int_{\eta^n_{i-1}}^{\eta^n_i}(\gamma-\gamma^n_i)^2\mu(d\gamma)\big)_{n\geq 1}$ due to Assumption 1. □

Lemma 3 (g^n satisfying Assumption 2). *Define* $g^n : [0, T] \mapsto \mathbb{R}$ *by* (12) *such that* $\theta : [0, T] \mapsto \mathbb{R}$ *satisfies* (5). *Under Assumption 1, for each* $\varepsilon > 0$, *there exists* $C_\varepsilon > 0$ *such that for any* $t, h \geq 0$ *with* $t + h \leq T$

$$\sup_{n\geq 1}|g^n(t) - g^n(t+h)| \leq C_\varepsilon h^{H-\varepsilon}.$$

Proof. Because θ satisfies (5), it is enough to prove that, for each fixed $\varepsilon > 0$, there exists $C > 0$ such that

$$\sup_{n\geq 1}\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|K^n(s)|ds \leq Ch^{H-\varepsilon}, \tag{40}$$

and

$$\sup_{n\geq 1}\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon}|K^n(s) - K^n(h+s)|ds \leq Ch^{H-\varepsilon}, \tag{41}$$

for any $t, h \geq 0$ with $t + h \leq T$. (40) being satisfied for the fractional kernel, it is enough to establish

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|K^n(s) - K(s)|ds \leq ch^{H-\varepsilon}.$$

In the proof of Lemma 1, it is shown that

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|K^n(s)-K(s)|ds$$

is bounded by (36), that is

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}\int_{\eta_n^n}^{\infty} e^{-\gamma s}\mu(d\gamma)ds + \sum_{i=1}^n \int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|c_i^n e^{-\gamma_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-\gamma s}\mu(d\gamma)|ds.$$

The first term is dominated by

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}\int_0^{\infty} e^{-\gamma s}\mu(d\gamma)ds = h^{H-\varepsilon}\frac{B(1/2-\varepsilon,H+1/2)}{B(1/2-H,H+1/2)},$$

where B is the usual Beta function. Moreover thanks to (16) and Assumption 1, we get

$$\sum_{i=1}^n \int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|c_i^n e^{-\gamma_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-\gamma s}\mu(d\gamma)|ds \le ch^{\frac{5}{2}-\varepsilon},$$

yielding (40). Similarly, we obtain (41) by showing that

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon}\left|(K^n(s)-\Delta_h K^n(s))-(K(s)-\Delta_h K(s))\right|ds \le ch^{H-\varepsilon}.$$

By similar computations as previously and using (39), we get that

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon}\left|(K^n(s)-\Delta_h K^n(s))-(K(s)-\Delta_h K(s))\right|ds$$

is dominated by

$$c\left(\int_0^t (t-s)^{\frac{1}{2}-\varepsilon}\int_{\eta_n^n}^{\infty}(1-e^{-\gamma h})e^{-\gamma s}\mu(d\gamma)ds + h\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n}(\gamma-\gamma_i^n)^2\mu(d\gamma)\right).$$

The first term being bounded by

$$\int_0^t (t-s)^{\frac{1}{2}-\varepsilon}\int_0^{\infty}(1-e^{-\gamma h})e^{-\gamma s}\mu(d\gamma)ds = \int_0^t (t-s)^{\frac{1}{2}-\varepsilon}(K(s)-K(h+s))ds \le ch^{H-\varepsilon},$$

Assumption 1 leads to (41).

□

5.3 Proof of Theorem 4

Uniform boundedness : We start by showing the uniform boundedness of the unique continuous solutions $(\psi^n(\cdot,a+ib))_{n\ge 1}$ of (25).

Proposition 5. *For a fixed $T>0$, there exists $C>0$ such that*

$$\sup_{n\geq 1}\sup_{t\in[0,T]}|\psi^n(t,a+ib)|\leq C\left(1+b^2\right),$$

for any $a\in[0,1]$ and $b\in\mathbb{R}$.

Proof. Let $z=a+ib$ and start by noticing that $\Re(\psi^n(\cdot,z))$ is non-positive because it solves the following linear Volterra equation with continuous coefficients

$$\chi=K^n*\left(f+\left(\rho\nu\Re(z)-\lambda+\frac{\nu^2}{2}\Re(\psi^n(\cdot,z))\right)\chi\right),$$

where

$$f=\frac{1}{2}\left(a^2-a-(1-\rho^2)b^2\right)-\frac{1}{2}(\rho b+\nu\psi^n(\cdot,z))^2$$

is continuous non-positive, see Theorem 7. In the same way $\Re(\psi(\cdot,z))$ is also non-positive. Moreover, observe that $\psi^n(\cdot,z)$ solves the following linear Volterra equation with continuous coefficients

$$\chi=K^n*\left(\frac{1}{2}(z^2-z)+(\rho\nu z-\lambda+\frac{\nu^2}{2}\psi^n(\cdot,z))\chi\right),$$

and

$$\Re\left(\rho\nu z-\lambda+\frac{\nu^2}{2}\psi^n(\cdot,z)\right)\leq\nu-\lambda.$$

Therefore, Corollary 4 leads to

$$\sup_{t\in[0,T]}|\psi^n(t,z)|\leq\frac{1}{2}|z^2-z|\int_0^T E^n_{\nu-\lambda}(s)ds,$$

where $E^n_{\nu-\lambda}$ denotes the canonical resolvent of K^n with parameter $\nu-\lambda$, see Appendix VIII.A.3. This resolvent converges in $\mathbb{L}^1([0,T],\mathbb{R})$ because K^n converges in $\mathbb{L}^1([0,T],\mathbb{R})$ to K, see [GLS90, Theorem 2.3.1]. Hence, $(\int_0^T E^n_{\nu-\lambda}(s)ds)_{n\geq 1}$ is bounded, which ends the proof. □

End of the proof of Theorem 4 : Set $z=a+ib$ and recall that

$$\psi^n(\cdot,z)=K^n*F(z,\psi^n(\cdot,z));\quad \psi(\cdot,z)=K*F(z,\psi(\cdot,z)).$$

with $F(z,x)=\frac{1}{2}\left(z^2-z\right)+(\rho\nu z-\lambda)x+\frac{\nu^2}{2}x^2$. Hence, for $t\in[0,T]$,

$$\psi(t,z)-\psi^n(t,z)=h^n(t,z)+K*\left(F(z,\psi(\cdot,z))-F(z,\psi^n(\cdot,z))\right)(t),$$

with $h^n(\cdot,z)=(K^n-K)*F(z,\psi^n(\cdot,z))$. Thanks to Proposition 5, we get the existence of a positive constant C such that

$$\sup_{n\geq 1}\sup_{t\in[0,T]}|h^n(t,a+ib)|\leq C(1+b^4)\int_0^T|K^n(s)-K(s)|ds, \tag{42}$$

for any $b \in \mathbb{R}$ and $a \in [0,1]$. Moreover notice that $(\psi - \psi^n - h^n)(\cdot, z)$ is solution of the following linear Volterra equation with continuous coefficients

$$\chi = K * \left(\left(\rho \nu z - \lambda + \frac{\nu^2}{2}(\psi + \psi_n)(\cdot, z) \right)(\chi + h^n(\cdot, z)) \right),$$

and remark that the real part of $\rho \nu z - \lambda + \frac{\nu^2}{2}(\psi + \psi_n)(\cdot, z)$ is dominated by $\nu - \lambda$ because $\Re(\psi(\cdot, z))$ and $\Re(\psi^n(\cdot, z))$ are non-positive. An application of Corollary 4 together with (42) ends the proof.

5.4 Proof of Proposition 3

We consider for each $n \geq 1$, (S^n, V^n) defined by the multi-factor Heston model in Definition 2 with $\sigma(x) = \nu\sqrt{x}$.

Tightness of $(\int_0^\cdot V_s^n ds, \int_0^\cdot \sqrt{V_s^n} dW_s, \int_0^\cdot \sqrt{V_s^n} dB_s)_{n \geq 1}$ **:** Because the process $\int_0^\cdot V_s^n ds$ is non-decreasing, it is enough to show that

$$\sup_{n \geq 1} \mathbb{E}[\int_0^T V_t^n dt] < \infty, \tag{43}$$

to obtain its tightness for the uniform topology. Recalling that $\sup_{t \in [0,T]} \mathbb{E}[V_t^n] < \infty$ from Proposition 7 in the Appendix, we get

$$\mathbb{E}\left[\int_0^t \sqrt{V_s^n} dB_s \right] = 0,$$

and then by Fubini theorem

$$E[V_t^n] = g^n(t) + \sum_{i=1}^n c_i^n \mathbb{E}[V_t^{n,i}],$$

with

$$\mathbb{E}[V_t^{n,i}] = \int_0^t (-\gamma_i^n \mathbb{E}[V_s^{n,i}] - \lambda \mathbb{E}[V_s^n]) ds.$$

Thus $t \mapsto \mathbb{E}[V_t^n]$ solves the following linear Volterra equation

$$\chi(t) = \int_0^t K^n(t-s) \left(-\lambda \chi(s) + \theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)} \right) ds,$$

with K^n given by (13). Theorem 5 in the Appendix leads to

$$\mathbb{E}[V_t^n] = \int_0^t E_\lambda^n(t-s) \left(\theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2} - H)} \right) ds,$$

and then by Fubini theorem again

$$\int_0^t \mathbb{E}[V_s^n] ds = \int_0^t \left(\int_0^{t-s} E_\lambda^n(u) du \right) \left(\theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2} - H)} \right) ds,$$

where E^n_λ is the canonical resolvent of K^n with parameter λ, defined in Appendix VIII.A.3. Because $(K^n)_{n\geq 1}$ converges to the fractional kernel K in $\mathbb{L}^1([0,T],\mathbb{R})$, we obtain the convergence of E^n_λ in $\mathbb{L}^1([0,T],\mathbb{R})$ to the canonical resolvent of K with parameter λ, see [GLS90, Theorem 2.3.1]. In particular thanks to Corollary 3 in the Appendix, $\int_0^t E^n_\lambda(s)ds$ is uniformly bounded in $t\in[0,T]$ and $n\geq 1$. This leads to (43) and then to the tightness of $(\int_0^\cdot V^n_s ds, \int_0^\cdot \sqrt{V^n_s}dW_s, \int_0^\cdot \sqrt{V^n_s}dB_s)_{n\geq 1}$ by [JS13, Theorem VI-4.13].

Convergence of $(S^n, \int_0^\cdot V^n_s ds)_{n\geq 1}$: We set $M^{n,1}_t = \int_0^t \sqrt{V^n_s}dW_s$ and $M^{n,2}_t = \int_0^t \sqrt{V^n_s}dB_s$. Denote by (X, M^1, M^2) a limit in law for the uniform topology of a subsequence of the tight family $(\int_0^\cdot V^n_s ds, M^{n,1}, M^{n,2})_{n\geq 1}$. An application of stochastic Fubini theorem, see [Ver12], yields

$$\int_0^t V^n_s ds = \int_0^t \int_0^{t-s} (K^n(u)-K(u))du dY^n_s + \int_0^t K(t-s)Y^n_s ds, \quad t\in[0,T], \tag{44}$$

where $Y^n_t = \int_0^t (s^{-H-\frac{1}{2}}\frac{V_0}{\Gamma(1/2-H)} + \theta(s) - \lambda V^n_s)ds + \nu M^{n,2}_t$. Because $(Y^n)_{n\geq 1}$ converges in law for the uniform topology to $Y=(Y_t)_{t\leq T}$ given by $Y_t = \int_0^t (s^{-H-\frac{1}{2}}\frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s))ds - \lambda X_t + \nu M^2_t$, we also get the convergence of $(\int_0^\cdot K(\cdot - s)Y^n_s ds)_{n\geq 1}$ to $\int_0^\cdot K(\cdot - s)Y_s ds$. Moreover, for any $t\in[0,T]$,

$$\left|\int_0^t \int_0^{t-s} (K^n(u)-K(u))du\left(s^{-H-\frac{1}{2}}\frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s) - \lambda V^n_s\right)ds\right|$$

is bounded by

$$\int_0^t |K^n(s)-K(s)|ds\left(\int_0^t (s^{-H-\frac{1}{2}}\frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s))ds + \lambda\int_0^t V^n_s ds\right),$$

which converges in law for the uniform topology to zero thanks to the convergence of $(\int_0^\cdot V^n_s ds)_{n\geq 1}$ together with Proposition 2. Finally,

$$\mathbb{E}\left[\left|\int_0^t \int_0^{t-s} (K^n(u)-K(u))du dM^{n,2}_s\right|^2\right] \leq c\int_0^T (K^n(s)-K(s))^2 ds \mathbb{E}\left[\int_0^t V^n_s ds\right],$$

which goes to zero thanks to (43) and Proposition 2. Hence, by passing to the limit in (44), we obtain

$$X_t = \int_0^t K(t-s)Y_s ds,$$

for any $t\in[0,T]$, almost surely. The processes being continuous, the equality holds on $[0,T]$. Then, by the stochastic Fubini theorem, we deduce that $X=\int_0^\cdot V_s ds$, where V is a continuous process defined by

$$V_t = \int_0^t K(t-s)dY_s = V_0 + \int_0^t K(t-s)(\theta(s)-\lambda V_s)ds + \nu\int_0^t K(t-s)dM^2_s.$$

Furthermore because $(M^{n,1}, M^{n,2})$ is a martingale with bracket

$$\int_0^{\cdot} V_s^n ds \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

[JS13, Theorem VI-6.26] implies that (M^1, M^2) is a local martingale with the following bracket

$$\int_0^{\cdot} V_s ds \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

By [RY13, Theorem V-3.9], there exists a two-dimensional Brownian motion $(\widetilde{W}, \widetilde{B})$ with $d\langle \widetilde{W}, \widetilde{B} \rangle_t = \rho dt$ such that

$$M_t^1 = \int_0^t \sqrt{V_s} d\widetilde{W}_s, \quad M_t^2 = \int_0^t \sqrt{V_s} d\widetilde{B}_s, \quad t \in [0, T].$$

In particular V is solution of the fractional stochastic integral equation in Definition 1 with $\sigma(x) = \nu\sqrt{x}$. Because $S^n = \exp(M^{n,1} - \frac{1}{2}\int_0^{\cdot} V_s^n ds)$, we deduce the convergence of $(S^n, \int_0^{\cdot} V_s^n ds)_{n\geq 1}$ to the limit point $(S, \int_0^{\cdot} V_s ds)$ that displays the rough-Heston dynamics of Definition 1. The uniqueness of such dynamics, see [AJLP17, MS15] and Chapter IV, enables us to conclude that $(S^n, V^n)_{n\geq 1}$ admits a unique limit point and hence converges to the rough Heston dynamics.

5.5 Proof of Proposition 4

We use the Lewis Fourier inversion method, see [Lew01], to write

$$C^n(k,T) - C(k,T) = S_0 \frac{e^{\frac{k}{2}}}{2\pi} \int_{b\in\mathbb{R}} \frac{e^{-ibk}}{b^2 + \frac{1}{4}} \Big(L(T, \frac{1}{2} + ib) - L^n(T, \frac{1}{2} + ib) \Big) db.$$

Hence,

$$|C^n(k,T) - C(k,T)| \leq S_0 \frac{e^{\frac{k}{2}}}{2\pi} \int_{b\in\mathbb{R}} \frac{1}{b^2 + \frac{1}{4}} \Big| L(T, \frac{1}{2} + ib) - L^n(T, \frac{1}{2} + ib) \Big| db. \tag{45}$$

Because $L(T,z)$ and $L^n(T,z)$ satisfy respectively the formulas (22) and (24) with g and g^n given by

$$g(t) = \int_0^t K(t-s)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)} + \theta(s)\big) ds, \quad g^n(t) = \int_0^t K^n(t-s)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)} + \theta(s)\big) ds,$$

and $\psi(\cdot, z)$ and $\psi^n(\cdot, z)$ solve respectively (23) and (25), we use the Fubini theorem to deduce that

$$L(T,z) = \exp\bigg(\int_0^T \psi(T-s, z)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)} + \theta(s)\big) ds \bigg), \tag{46}$$

and

$$L^n(T,z) = \exp\bigg(\int_0^T \psi^n(T-s, z)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)} + \theta(s)\big) ds \bigg), \tag{47}$$

with $z = 1/2 + ib$. Therefore, relying on the local Lipschitz property of the exponential function, it suffices to find an upper bound for $\Re(\psi^n(\cdot, z))$ in order to get an error for the price of the call from (45). This is the object of the next paragraph.

Upper bound of $\Re(\psi^n(\cdot,z))$: We denote by $\phi^n_\eta(\cdot,b)$ the unique continuous function satisfying the following Riccati Volterra equation

$$\phi^n_\eta(\cdot,b) = K^n * \left(-b + \eta\phi^n_\eta(\cdot,b) + \frac{\nu^2}{2}\phi^n_\eta(\cdot,b)^2\right),$$

with $b \geq 0$ and $\eta, \nu \in \mathbb{R}$.

Proposition 6. *Fix $b_0, t_0 \geq 0$ and $\eta \in \mathbb{R}$. The functions $b \mapsto \phi^n_\eta(t_0,b)$ and $t \mapsto \phi^n_\eta(t,b_0)$ are non-increasing on $\mathbb{R}_+$. Furthermore*

$$\phi^n_\eta(t,b) \leq \frac{1-\sqrt{1+2b\nu^2(\int_0^t E^n_\eta(s)ds)^2}}{\nu^2\int_0^t E^n_\eta(s)ds}, \quad t>0,$$

where E^n_η is the canonical resolvent of K^n with parameter η defined in Appendix VIII.A.3.

Proof. The claimed monotonicity of $b \mapsto \phi^n_\eta(t_0,b)$ is directly obtained from Theorem 7. Consider now $h, b_0 > 0$. It is easy to see that $\Delta_h\phi^n_\eta(\cdot,b_0)$ solves the following Volterra equation

$$\Delta_h\phi^n_\eta(b_0,t) = \left(\Delta_t K^n * F(\phi^n_\eta(\cdot,b_0))\right)(h) + \left(K^n * F(\Delta_h\phi^n_\eta(\cdot,b_0))\right)(t)$$

with $F(b,x) = -b + \eta x + \frac{\nu^2}{2}x^2$. Notice that $t \to -\left(\Delta_t K^n * F(\phi^n_\eta(\cdot,b_0))\right)(h) \in \mathcal{G}_K$, defined in Appendix VIII.C, thanks to Theorem 7. $\phi^n_\eta(\cdot,b) - \Delta_h\phi^n_\eta(\cdot,b)$ being solution of the following linear Volterra integral equation with continuous coefficients,

$$x(t) = -\left(\Delta_t K^n * F(b,\phi^n_\eta(\cdot,b_0))\right)(h) + \left(K^n * \left(\left(\eta + \frac{\nu^2}{2}(\phi^n_\eta(\cdot,b) + \Delta_h\phi^n_\eta(\cdot,b))\right)x\right)\right)(t),$$

we deduce its non-negativity using again Theorem 7. Thus, $t \in \mathbb{R}_+ \to \phi^n_\eta(t,b_0)$ is non-increasing and consequently $\sup_{s\in[0,t]}|\phi_\eta(s,b)| = |\phi^n_\eta(t,b_0)|$ as $\phi^n_\eta(0,b) = 0$. Hence, Theorem 5 leads to

$$\phi^n_\eta(t,b) = \int_0^t E^n_\eta(t-s)(-b + \frac{\nu^2}{2}\phi^n_\eta(s,b)^2) \leq \int_0^t E^n_\eta(s)ds\left(-b + \frac{\nu^2}{2}\phi^n_\eta(t,b)^2\right).$$

We end the proof by solving this inequality of second order in $\phi^n_\eta(t,b)$ and using that ϕ^n_η is non-positive. Notice that $\int_0^t E^n_\eta(s)ds > 0$ for each $t > 0$, see Corollary 3. □

Corollary 1. *Fix $a \in [0,1]$. We have, for any $t \in (0,T]$ and $b \in \mathbb{R}$,*

$$\sup_{n\geq 1}\Re(\psi^n(t,a+ib)) \leq \frac{1-\sqrt{1+(a-a^2+(1-\rho^2)b^2)\nu^2 m(t)^2}}{\nu^2 m(t)}$$

where $m(t) = \inf_{n\geq 1}\int_0^t E^n_{\rho\nu a-\lambda}(s)ds > 0$ for all $t \in (0,T]$ and E^n_η is the canonical resolvent of K^n with parameter η defined in Appendix VIII.A.3.

Proof. Let $r = a - a^2 + (1-\rho^2)b^2$ and $\eta = \rho\nu a - \lambda$. $\phi_\eta^n(\cdot, r) - \Re(\psi^n(\cdot, a+ib))$ being solution of the following linear Volterra equation with continuous coefficients

$$\chi = K * \left(\frac{1}{2}\left(\rho b + \nu\Im(\psi^n(\cdot, a+ib))\right)^2 + \left(\rho\nu a - \lambda + \frac{\nu^2}{2}\left(\Re(\psi^n(\cdot, a+ib)) + \phi_\eta(\cdot, r)\right)\right)\chi\right),$$

we use Theorem 7 together with Proposition 6 to get, for all $t \in [0,T]$ and $b \in \mathbb{R}$,

$$\Re(\psi^n(t, a+ib)) \le \frac{1 - \sqrt{1 + 2r\nu^2(\int_0^t E_\eta^n(s)ds)^2}}{\nu^2 \int_0^t E_\eta^n(s)ds}. \tag{48}$$

Moreover for any $t \in [0,T]$, $\int_0^t E_\eta^n(s)ds$ converges as n goes to infinity to $\int_0^t E_\eta(s)ds$ because K^n converges to K in $\mathbb{L}^1([0,T],\mathbb{R})$, see [GLS90, Theorem 2.3.1], where E_η denotes the canonical resolvent of K with parameter η. Therefore, $m(t) = \inf_{n\ge1}\int_0^t E_\eta^n(s)ds > 0$, for all $t \in (0,T]$, because $\int_0^t E_\eta(s)ds > 0$ and $\int_0^t E_\eta^n(s)ds > 0$ for all $n \ge 1$, see Corollary 3. Finally we end the proof by using (48) together with the fact that $x \mapsto \frac{1-\sqrt{1+2r\nu^2x^2}}{\nu^2 x}$ is non-increasing on $(0,\infty)$.

End of the proof of Proposition 4 : Assume that $|\rho| < 1$ and fix $a = 1/2$. By dominated convergence theorem,

$$\int_0^T \frac{1 - \sqrt{1 + (a - a^2 + (1-\rho^2)b^2)\nu^2 m(T-s)^2}}{\nu^2 m(T-s)}(\theta(s) + V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds$$

is equivalent to

$$-|b|\frac{\sqrt{1-\rho^2}}{\nu}\int_0^T (\theta(s) + V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds,$$

as b tends to infinity. Hence, thanks to Corollary 1, there exists $C > 0$ such that for any $b \in \mathbb{R}$

$$\sup_{n\ge1} \Re(\psi^n(t, a+ib)) \le C(1 - |b|). \tag{49}$$

Recalling that

$$\forall z_1, z_2 \in \mathbb{C} \text{ such that } \Re(z_1), \Re(z_2) \le c, \quad |e^{z_1} - e^{z_2}| \le e^c|z_1 - z_2|,$$

we obtain

$$|L^n(a+ib,T) - L(a+ib,T)| \le e^{C(1-|b|)}\sup_{t\in[0,T]}|\psi^n(t,a+ib) - \psi(t,a+ib)|\int_0^T (\theta(s) + V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds,$$

from (46), (47) and (49). We deduce Proposition 4 thanks to (45) and Theorem 4 together with the fact that $\int_{b\in\mathbb{R}}\frac{b^4+1}{b^2+\frac{1}{4}}e^{C(1-|b|)}db < \infty$. □

VIII.A Stochastic convolutions and resolvents

We recall in this Appendix the framework and notations introduced in [AJLP17].

VIII.A.1 Convolution notation

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 by

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

whenever these expressions are well-defined. If F is a function on $\mathbb{R}_+$, we write $K * F = K * (F dt)$, that is

$$(K * F)(t) = \int_0^t K(t-s)F(s)ds.$$

We can show that $L * F$ is almost everywhere well-defined and belongs to $\mathbb{L}^p_{loc}(\mathbb{R}_+, \mathbb{R})$, whenever $F \in \mathbb{L}^p_{\text{loc}}(\mathbb{R}_+, \mathbb{R})$. Moreover, $(F * G) * L = F * (G * L)$ *a.e.*, whenever $F, G \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$, see [GLS90, Theorem 3.6.1 and Corollary 3.6.2] for further details.

For any continuous semimartingale $M = \int_0^\cdot b_s ds + \int_0^\cdot a_s dB_s$ the convolution

$$(K * dM)_t = \int_0^t K(t-s)dM_s$$

is well-defined as an Itô integral for every $t \geq 0$ such that

$$\int_0^t |K(t-s)||b_s|ds + \int_0^t |K(t-s)|^2|a_s|^2 ds < \infty.$$

Using stochastic Fubini Theorem, see [AJLP17, Lemma 2.1], we can show that for each $t \geq 0$, almost surely

$$(L * (K * dM))_t = ((L * K) * dM)_t, \tag{50}$$

whenever $K \in \mathbb{L}^2_{loc}(\mathbb{R}_+, \mathbb{R})$ and a, b are locally bounded *a.s.*

Finally from Lemma 2.4 in [AJLP17] together with the Kolmogorov continuity theorem, we can show that there exists a unique version of $(K * dM_t)_{t\geq 0}$ that is continuous whenever b and σ are locally bounded. In this paper, we will always work with this continuous version.

Note that the convolution notation could be easily extended for matrix-valued K and L. In this case, the associativity properties exposed above hold.

VIII.A.2 Resolvent of the first kind

We define the *resolvent of the first kind* of a $d \times d$-matrix valued kernel K, as the $\mathbb{R}^{d\times d}$-valued measure L on $\mathbb{R}_+$ of locally bounded variation such that

$$K * L = L * K \equiv \text{id},$$

where id stands for the identity matrix, see [GLS90, Definition 5.5.1]. The resolvent of the first kind does not always exist. In the case of the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ the resolvent of the first kind exists and is given by

$$L(dt) = \frac{t^{-H-\frac{1}{2}}}{\Gamma(1/2-H)}dt,$$

for any $H \in (0,1/2)$. If K is non-negative, non-increasing and not identically equal to zero on $\mathbb{R}_+$, the existence of a resolvent of the first kind is guaranteed by [GLS90, Theorem 5.5.5].

The following result shown in [AJLP17, Lemma 2.6], is stated here for $d = 1$ but is true for any dimension $d \geq 1$.

Lemma 4. *Assume that $K \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$ admits a resolvent of first kind L. For any $F \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$ such that $F * L$ is right-continuous and of locally bounded variation one has*

$$F = (F * L)(0)K + d(F * L) * K.$$

Here, df denotes the measure such that $f(t) = f(0) + \int_{[0,t]} df(s)$, for all $t \geq 0$, for any right-continuous function of locally bounde variation f on $\mathbb{R}_+$.

Remark 2. *The previous lemma will be used with $F = \Delta_h K$, for fixed $h > 0$. If K is continuous on $(0,\infty)$, then $\Delta_h K * L$ is right-continuous. Moreover, if K is non-negative and L non-increasing in the sense that $s \to L([s, s+t])$ is non-increasing for all $t \geq 0$, then $\Delta_h K * L$ is non-decreasing since*

$$(\Delta_h K * L)(t) = 1 - \int_{[0,h)} K(h-s)L(t+ds), \qquad t > 0.$$

In particular, $\Delta_h K * L$ is of locally bounded variation.

VIII.A.3 Resolvent of the second kind

We consider a kernel $K \in \mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ and define *the resolvent of the second kind of K* as the unique function $R_K \in \mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ such that

$$K - R_K = K * R_K.$$

For $\lambda \in \mathbb{R}$, we define *the canonical resolvent of K with parameter λ* as the unique solution $E_\lambda \in \mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ of

$$E_\lambda - K = \lambda K * E_\lambda.$$

This means that $E_\lambda = -R_{-\lambda K}/\lambda$, when $\lambda \neq 0$ and $E_0 = K$. The existence and uniqueness of R_K and E_λ is ensured by [GLS90, Theorem 2.3.1] together with the continuity of $K \to E_\lambda(K)$ in the topology of $\mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$. Moreover, if $K \in \mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ so does E_λ due to [GLS90, Theorem 2.3.5].

We recall [GLS90, Theorem 2.3.5] regarding the existence and uniqueness of a solution of linear Volterra integral equations in $\mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$.

Theorem 5. *Let $f \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$. The integral equation*

$$x = f + \lambda K * x$$

admits a unique solution $x \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$ given by

$$x = f + \lambda E_\lambda * f.$$

When K and λ are positive, E_λ is also positive, see [GLS90, Proposition 9.8.1]. In that case, we have a Grönwall type inequality given by [GLS90, Lemma 9.8.2].

Lemma 5. *Let $x, f \in \mathbb{L}^1_{loc}(\mathbb{R}_+, \mathbb{R})$ such that*

$$x(t) \leq (\lambda K * x)(t) + f(t), \quad t \geq 0, \ a.e.$$

Then,

$$x(t) \leq f(t) + (\lambda E_\lambda * f)(t), \quad t \geq 0, \ a.e.$$

Note that the definition of the resolvent of the second kind and canonical resolvent can be extended for matrix-valued kernels. In that case, Theorem 5 still holds.

Remark 3. *The canonical resolvent of the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ with parameter λ is given by*

$$t^{\alpha-1} E_\alpha(-\lambda t^\alpha),$$

where $E_\alpha(x) = \sum_{k\geq 0} \frac{x^k}{\Gamma(\alpha(k+1))}$ is the Mittag-Leffler function and $\alpha = H + 1/2$ for $H \in (0, 1/2)$.

VIII.B Some existence results for stochastic Volterra equations

We collect in this Appendix existence results for general stochastic Volterra equations as introduced in [AJLP17]. We refer to [AJLP17] and Chapter IV for the proofs. We fix $T > 0$ and consider the d-dimensional stochastic Volterra equation

$$X_t = g(t) + \int_0^t K(t-s) b(X_s) ds + \int_0^t K(t-s) \sigma(X_s) dB_s, \quad t \in [0, T], \tag{51}$$

where $b : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ are continuous functions with linear growth, $K \in \mathbb{L}^2([0, T], \mathbb{R}^{d\times d})$ is a kernel admitting a resolvent of the first kind L, $g : [0, T] \mapsto \mathbb{R}^d$ is a continuous function and B is a m-dimensional Brownian motion on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$. In order to prove the weak existence of continuous solutions to (51), the following regularity assumption is needed.

Assumption 3. *There exists $\gamma > 0$ and $C > 0$ such that for any $t, h \geq 0$ with $t + h \leq T$,*

$$|g(t+h) - g(t)|^2 + \int_0^h |K(s)|^2 ds + \int_0^{T-h} |K(h+s) - K(s)|^2 ds \leq C h^{2\gamma}.$$

The following existence result can be found in Theorem 6 of Chapter IV.

Proposition 7. *Under Assumption 3, the stochastic Volterra equation* (51) *admits a weak continuous solution* $X = (X_t)_{t \le T}$. *Moreover* X *satisfies*

$$\sup_{t \in [0,T]} \mathbb{E}[|X_t|^p] < \infty, \quad p > 0, \tag{52}$$

and admits Hölder continuous paths on $[0,T]$ *of any order strictly less than* γ.

In particular, for the fractional kernel, Proposition 7 yields the following result.

Corollary 2. *Fix* $H \in (0, 1/2)$ *and* $\theta : [0,T] \mapsto \mathbb{R}$ *satisfying*

$$\forall \varepsilon > 0, \quad \exists C_\varepsilon > 0; \quad \forall u \in (0,T] \quad |\theta(u)| \le C_\varepsilon u^{-\frac{1}{2}-\varepsilon}.$$

The fractional stochastic integral equation

$$X_t = X_0 + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-u)^{H-\frac{1}{2}} (\theta(u) + b(X_u)) du + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-u)^{H-\frac{1}{2}} \sigma(X_u) dB_u,$$

admits a weak continuous solution $X = (X_t)_{t \le T}$ *for any* $X_0 \in \mathbb{R}$. *Moreover* X *satisfies* (52) *and admits Hölder continuous paths on* $[0,T]$ *of any order strictly less than* H.

Proof. It is enough to notice that the fractional stochastic integral equation is a particular case of (51) with $d = m = 1$, $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ the fractional kernel, which admits a resolvent of the first kind, see Section VIII.A.2, and

$$g(t) = X_0 + \frac{1}{\Gamma(1/2+H)} \int_0^t (t-u)^{H-1/2} \theta(u) du.$$

As $t \mapsto t^{1/2+\varepsilon}\theta(t)$ is bounded on $[0,T]$, we may show that g is $H - \varepsilon$ Hölder continuous for any $\varepsilon > 0$. Hence, Assumption 3 is satisfied and the claimed result is directly obtained from Proposition 7. □

We now establish the strong existence and uniqueness of (51) in the particular case of smooth kernels. This is done by extending the Yamada-Watanabe pathwise uniqueness proof in [YW71].

Proposition 8. *Fix* $m = d = 1$ *and assume that* g *is Hölder continuous,* $K \in C^1([0,T], \mathbb{R})$ *admitting a resolvent of the first kind and that there exists* $C > 0$ *and* $\eta \in [1/2, 1]$ *such that for any* $x, y \in \mathbb{R}$,

$$|b(x) - b(y)| \le C|x - y|, \quad |\sigma(x) - \sigma(y)| \le C|x - y|^\eta.$$

Then, the stochastic Volterra equation (51) *admits a unique strong continuous solution.*

Proof. We start by noticing that, K being smooth, it satisfies Assumption 3. Hence, the existence of a weak continuous solution to (51) follows from Proposition 7. It is therefore enough to show the pathwise uniqueness. We may proceed similarly to [YW71] by considering $a_0 = 1$, $a_{k-1} > a_k$ for $k \ge 1$ with $\int_{a_k}^{a_{k-1}} x^{-2\eta} dx = k$ and $\varphi_k \in C^2(\mathbb{R}, \mathbb{R})$ such that $\varphi_k(x) = \varphi_k(-x)$, $\varphi_k(0) = 0$ and for $x > 0$

- $\varphi_k'(x) = 0$ for $x \leq a_k$, $\varphi_k'(x) = 1$ for $x \geq a_{k-1}$ and $\varphi_k'(x) \in [0,1]$ for $a_k < x < a_{k-1}$.
- $\varphi_k''(x) \in [0, \frac{2}{k} x^{-2\eta}]$ for $a_k < x < a_{k-1}$.

Let X^1 and X^2 be two solutions of (51) driven by the same Brownian motion B. Notice that, thanks to the smoothness of K, $X^i - g$ are semimartingales and for $i = 1,2$

$$d(X_t^i - g(t)) = K(0) dY_t^i + (K' * dY^i)_t \, dt,$$

with $Y_t^i = \int_0^t b(X_s^i) ds + \int_0^t \sigma(X_s^i) dB_s$. Using Itô's formula, we write

$$\varphi_k(X_t^2 - X_t^1) = I_t^1 + I_t^2 + I_t^3,$$

where

$$I_t^1 = K(0) \int_0^t \varphi_k'(X_s^2 - X_s^1) d(Y_s^1 - Y_s^2),$$

$$I_t^2 = \int_0^t \varphi_k'(X_s^2 - X_s^1)(K' * d(Y^1 - Y^2))_s ds,$$

$$I_t^3 = \frac{K(0)^2}{2} \int_0^t \varphi_k''(X_s^2 - X_s^1)(\sigma(X_s^2) - \sigma(X_s^1))^2 ds.$$

Recalling that $\sup_{t \leq T} \mathbb{E}[(X_t^i)^2] < \infty$ for $i = 1,2$ from Proposition 7, we obtain that

$$\mathbb{E}[I_t^1] \leq \mathbb{E}[K(0) \int_0^t |b(X_s^2) - b(X_s^1)| ds] \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|] ds,$$

and

$$\mathbb{E}[I_t^2] \leq c \int_0^t \mathbb{E}[(|K'| * |b(X^2) - b(X^1)|)_s] ds \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|] ds,$$

because b is Lipschitz continuous and K' is bounded on $[0,T]$. Finally by definition of φ_k and the η-Hölder continuity of σ, we have

$$\mathbb{E}[I_t^3] \leq \frac{c}{k},$$

which goes to zero when k is large. Moreover $\mathbb{E}[\varphi_k(X_t^2 - X_t^1)]$ converges to $\mathbb{E}[|X_t^2 - X_t^1|]$ when k tends to infinity, thanks to the monotone convergence theorem. Thus, we pass to the limit and obtain

$$\mathbb{E}[|X_t^2 - X_t^1|] \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|] ds.$$

Grönwall's lemma leads to $\mathbb{E}[|X_t^2 - X_t^1|] = 0$ yielding the claimed pathwise uniqueness.

□

Under additional conditions on g and K one can obtain the existence of non-negative solutions to (51) in the case of $d = m = 1$. As in [AJLP17, Theorem 3.5], the following assumption is needed.

Assumption 4. *We assume that $K \in \mathbb{L}^2([0,T],\mathbb{R})$ 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 $0 \le L([s,s+t]) \le L([0,t])$ for all $s,t \ge 0$ with $s+t \le T$.*

In Chapter IV, the proof of [AJLP17, Theorem 3.5] is adapted to prove the existence of a non-negative solution for a wide class of admissible input curves g satisfying[6]

$$\Delta_h g - (\Delta_h K * L)(0) g - d(\Delta_h K * L) * g \ge 0, \quad h \ge 0. \tag{53}$$

We therefore define the following set of admissible input curves

$$\mathcal{G}_K = \left\{ g : [0,T] \mapsto \mathbb{R} \text{ continuous satisfying (53) and } g(0) \ge 0 \right\}.$$

The following existence theorem is a particular case of Theorem 7 in Chapter IV.

Theorem 6. *Assume that $d = m = 1$ and that b and σ satisfy the boundary conditions*

$$b(0) \ge 0, \quad \sigma(0) = 0.$$

Then, under Assumptions 3, and 4, the stochastic Volterra equation (51) admits a non-negative weak solution for any $g \in \mathcal{G}_K$.

Remark 4. *Note that any locally square-integrable completely monotone kernel [7] that is not identically zero satisfies Assumption 4, see [AJLP17, Example 3.6]. In particular, this is the case for*

- *the fractional kernel $K(t) = \frac{t^{H-1/2}}{\Gamma(H+1/2)}$, with $H \in (0,1/2)$.*
- *any weighted sum of exponentials $K(t) = \sum_{i=1}^n c_i e^{-\gamma_i t}$ such that $c_i, \gamma_i \ge 0$ for all $i \in \{1,\dots,n\}$ and $c_i > 0$ for some i.*

Remark 5. *Theorem 6 will be used with functions g of the following form*

$$g(t) = c + \int_0^t K(t-s)\xi(ds),$$

where ξ is a non-negative measure of locally bounded variation and c is a non-negative constant. In that case, we may show that (53) is satisfied, under Assumption 4.

[6] Under Assumption 4 one can show that $\Delta_h K * L$ is non-increasing and right-continuous thanks to Remark 2 so that the associated measure $d(\Delta_h K * L)$ is well-defined.

[7] A kernel $K \in \mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R})$ is said to be completely monotone, if it is infinitely differentiable on $(0,\infty)$ such that $(-1)^j K^{(j)}(t) \ge 0$ for any $t > 0$ and $j \ge 0$.

VIII.C Linear Volterra equation with continuous coefficients

In this section, we consider $K \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ satisfying Assumption 4 with $T=\infty$ and recall the definition of $\mathscr{G}_K$, that is

$$\mathscr{G}_K = \left\{g : \mathbb{R}_+ \mapsto \mathbb{R} \text{ continuous satisfying (53) and } g(0) \geq 0\right\}.$$

We denote by $\|.\|_{\infty,T}$ the usual uniform norm on $[0,T]$, for each $T>0$.

Theorem 7. *Let $K \in \mathbb{L}^2_{loc}(\mathbb{R}_+,\mathbb{R})$ satisfying Assumption 4 and $g,z,w : \mathbb{R}_+ \mapsto \mathbb{R}$ be continuous functions. The linear Volterra equation*

$$\chi = g + K * \left(z\chi + w\right) \tag{54}$$

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

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

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

Proof. The existence and uniqueness of such solution in $\chi \in \mathbb{L}^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ is obtained from [AJLP17, Lemma C.1]. Because χ is solution of (54), it is enough to show the local boundedness of χ to get its continuity. This follows from Grönwall's Lemma 5 applied on the following inequality

$$|\chi(t)| \leq \|g\|_{\infty,T} + \left(K * (\|z\|_{\infty,T}|\chi|(.) + \|w\|_{\infty,T})\right)(t),$$

for any $t \in [0,T]$ and for a fixed $T>0$.

We assume now that $g \in \mathscr{G}_K$ and w is non-negative. The fact that $g_{t_0} \in \mathscr{G}_K$, for $t_0 \geq 0$, is proved by adapting the computations of the proof of Theorem 3 in Chapter IV with $\nu = 0$ provided that χ is non-negative. In order to establish the non-negativity of χ, we introduce, for each $\varepsilon > 0$, χ_ε as the unique continuous solution of

$$\chi_\varepsilon = g + K * \left(z\chi_\varepsilon + w + \varepsilon\right). \tag{55}$$

It is enough to prove that χ_ε is non-negative, for every $\varepsilon > 0$, and that $(\chi_\varepsilon)_{\varepsilon>0}$ converges uniformly on every compact to χ as ε goes to zero.

Positivity of χ_ε : It is easy to see that χ_ε is non-negative on a neighborhood of zero because, for small t,

$$\chi_\varepsilon(t) = g(t) + \left(z(0)g(0) + w(0) + \varepsilon\right)\int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

as χ, z and w are continuous functions. Hence, $t_0 = \inf\{t>0; \quad \chi_\varepsilon(t) < 0\}$ is positive. If we assume that $t_0 < \infty$, we get $\chi_\varepsilon(t_0) = 0$ by continuity of χ_ε. χ_ε being the solution of (55), we have

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

with $g_{t_0,\varepsilon}(t) = \Delta_{t_0} g(t) + (\Delta_t K * (z\chi_\varepsilon + w + \varepsilon))(t_0)$. Then, by using Lemma 4 with $F = \Delta_t K$, we obtain

$$g_{t_0,\varepsilon}(t) = \Delta_{t_0} g(t) - (d(\Delta_t K * L) * g)(t_0) - (\Delta_t K * L)(0) g(t_0) \\ + (d(\Delta_t K * L) * \chi_\varepsilon)(t_0) + (\Delta_t K * L)(0)\chi_\varepsilon(t_0),$$

which is continuous and non-negative, because $g \in \mathscr{G}_K$ and $\Delta_t K * L$ is non-decreasing for any $t \geq 0$, see Remark 2. Hence, in the same way, $\Delta_{t_0}\chi_\varepsilon$ is non-negative on a neighborhood of zero. Thus $t_0 = \infty$, which means that χ_ε is non-negative.

Uniform convergence of χ_ε : We use the following inequality

$$|\chi - \chi_\varepsilon|(t) \leq \big(K * (\|z\|_{\infty,T}|\chi - \chi_\varepsilon| + \varepsilon)\big)(t), \quad t \in [0, T],$$

together with the Gronwall Lemma 5 to show the uniform convergence on $[0, T]$ of χ_ε to χ as ε goes to zero. In particular, χ is also non-negative.

□

Corollary 3. *Let $K \in \mathbb{L}^2_{loc}(\mathbb{R}_+,\mathbb{R})$ satisfying Assumption 4 and define E_λ as the canonical resolvent of K with parameter $\lambda \in \mathbb{R} - \{0\}$. Then, $t \mapsto \int_0^t E_\lambda(s)ds$ is non-negative and non-decreasing on $\mathbb{R}_+$. Furthermore $\int_0^t E_\lambda(s)ds$ is positive, if K does not vanish on $[0, t]$*

Proof. The non-negativity of $\chi = \int_0^\cdot E_\lambda(s)ds$ is obtained from Theorem 7 and from the fact that χ is solution of the following linear Volterra equation

$$\chi = K * (\lambda\chi + 1),$$

by Theorem 5. For fixed $t_0 > 0$, $\Delta_{t_0}\chi$ satisfies

$$\Delta_{t_0}\chi = g_{t_0} + K * (\lambda\Delta_{t_0}\chi + 1),$$

with $g_{t_0}(t) = \big(\Delta_t K * (\lambda\Delta_{t_0}\chi + 1)\big)(t_0) \in \mathscr{G}_K$, see Theorem 7. It follows that $\Delta_{t_0}\chi - \chi$ solves

$$x = g_{t_0} + K * (\lambda x).$$

Hence, another application of Theorem 7 yields that $\chi \leq \Delta_{t_0}\chi$, proving that $t \to \int_0^t E_\lambda(s)ds$ is non-decreasing. □

We now provide a version of Theorem 7 for complex valued solutions.

Theorem 8. *Let $z, w : \mathbb{R}_+ \mapsto \mathbb{C}$ be continuous functions and $h_0 \in \mathbb{C}$. The following linear Volterra equation*

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

admits unique continuous solution $h : \mathbb{R}_+ \mapsto \mathbb{C}$ such that

$$|h(t)| \leq \psi(t), \quad t \geq 0,$$

where $\psi : \mathbb{R}_+ \mapsto \mathbb{R}$ is the unique continuous solution of

$$\psi = |h_0| + K * (\Re(z)\psi + |w|).$$

Proof. The existence and uniqueness of a continuous solution is obtained in the same way as in the proof of Theorem 7. Consider now, for each $\varepsilon > 0$, ψ_ε the unique continuous solution of

$$\psi_\varepsilon = |h_0| + K * (\Re(z)\psi + |w| + \varepsilon).$$

As done in the proof of Theorem 7, ψ_ε converges uniformly on every compact to ψ as ε goes to zero. Thus, it is enough to show that, for every $\varepsilon > 0$ and $t \geq 0$,

$$|h(t)| \leq \psi_\varepsilon(t).$$

We start by showing the inequality in a neighborhood of zero. Because z, h, w and ψ_ε are continuous, we get, taking $h_0 = 0$,

$$|h(t)| = |w(0)| \int_0^t K(s)ds + o(\int_0^t K(s)ds), \quad \psi_\varepsilon(t) = (|w(0)| + \varepsilon) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

for small t. Hence, $|h| \leq \psi_\varepsilon$ on a neighborhood of zero. This result still holds when h_0 is not zero. Indeed in that case, it is easy to show that for t going to zero,

$$|h(t)|^2 = |h_0|^2 + 2\Re\big(\overline{h_0}(z(0)h_0 + w(0))\big) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and

$$|\psi_\varepsilon(t)|^2 = |h_0|^2 + 2\big(\Re(z(0))|h_0|^2 + |w(0)||h_0| + \varepsilon|h_0|\big) \int_0^t K(s)ds + o(\int_0^t K(s)ds).$$

As $|h_0|$ is now positive, we conclude that $|h| \leq \psi_\varepsilon$ on a neighborhood of zero by the Cauchy-Schwarz inequality.

Hence, $t_0 = \inf\{t > 0; \quad \psi_\varepsilon(t) < |h(t)|\}$ is positive. If we assume that $t_0 < \infty$, we would get that $|h(t_0)| = \psi_\varepsilon(t_0)$ by continuity of h and ψ_ε. Moreover,

$$\Delta_{t_0} h = \phi_h + K * (\Delta_{t_0} z \Delta_{t_0} h + \Delta_{t_0} w),$$

and

$$\Delta_{t_0} \psi_\varepsilon = \phi_{\psi_\varepsilon} + K * (\Delta_{t_0} \Re(z) \Delta_{t_0} w + \Delta_{t_0} |w| + \varepsilon).$$

An application of Lemma 4 with $F = \Delta_t K$ for $t > 0$, yields

$$\phi_h(t) = h_0(1 - (\Delta_t K * L)(t_0)) + (d(\Delta_t K * L) * h)(t_0) + (\Delta_t K * L)(0) h(t_0),$$

and

$$\phi_{\psi_\varepsilon}(t) = |h_0|(1 - (\Delta_t K * L)(t_0)) + (d(\Delta_t K * L) * \psi_\varepsilon)(t_0) + (\Delta_t K * L)(0)|h(t_0)|.$$

Relying on the fact that $d(\Delta_t K * L)$ is a non-negative measure and $\Delta_t K * L \leq 1$, by Remark 2, together with the fact that $|h(s)| \leq \psi_\varepsilon(s)$ for $s \leq t_0$, we get that $|\phi_h(t)| \leq \phi_{\psi_\varepsilon}(t)$. We now notice that in the case $h(t_0) = 0$, we have

$$\Delta_{t_0} h(t) = \phi_h(t) + w(t_0) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and

$$\Delta_{t_0}\psi_\varepsilon(t) = \phi_{\psi_\varepsilon}(t) + (|w(t_0)| + \varepsilon)\int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and in the case $|h(t_0)| > 0$, we have

$$\begin{aligned}|\Delta_{t_0}h(t)|^2 &= 2\big(\Re(z(t_0))|h(t_0)|^2 + \Re(w(t_0))\Re(h(t_0)) + \Im(w(t_0))\Im(h(t_0))\big)\int_0^t K(s)ds\\&\quad + |\phi_h(t)|^2 + o(\int_0^t K(s)ds),\\ \Delta_{t_0}\psi_\varepsilon(t)^2 &= 2\big(\Re(z(t_0))|h(t_0)|^2 + |w(t_0)||h(t_0)| + \varepsilon|h(t_0)|)\big)\int_0^t K(s)ds\\&\quad + \phi_{\psi_\varepsilon}(t)^2 + o(\int_0^t K(s)ds),\end{aligned}$$

for small t, thanks to the continuity of $z, w, h, \phi_h, \phi_{\psi_\varepsilon}$ and ψ_ε. In both cases, we obtain that $|h| \le \psi_\varepsilon$ on a neighborhood of t_0. Therefore $t_0 = \infty$ and for any $t \ge 0$

$$|h(t)| \le \psi_\varepsilon(t).$$

□

The following result is a direct consequence of Theorems 7 and 8.

Corollary 4. *Let $h_0 \in \mathbb{C}$ and $z, w : \mathbb{R}_+ \to \mathbb{C}$ be continuous functions such that $\Re(z) \le \lambda$ 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 $t \in [0, T]$,

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

where E_λ is the canonical resolvent of K with parameter λ.

Proof. From Theorem 8, we obtain that $|h| \le \psi_1$, where ψ_1 is the unique continuous solution of

$$\psi_1 = |h_0| + K * (\Re(z)\psi_1 + |w|).$$

Moreover define ψ_2 as the unique continuous solution of

$$\psi_2 = |h_0| + K * (\lambda\psi_2 + \|w\|_{\infty,T}).$$

Then, $\psi_2 - \psi_1$ solves

$$\chi = K * (\lambda\chi + f),$$

with $f = (\lambda - \Re(z))\psi_1 + \|w\|_{\infty,T} - w$, which is a non-negative function on $[0, T]$. Theorem 7 now yields

$$|h| \le \psi_1 \le \psi_2.$$

Finally, the claimed bound follows by noticing that, for $t \in [0, T]$,

$$\psi_2(t) = |h_0| + (\|w\|_{\infty,T} + \lambda|h_0|) \int_0^t E_\lambda(s) ds,$$

by Theorem 5 and that $\int_0^\cdot E_\lambda(s) ds$ is non-decreasing by Corollary 3. □

Part VII

Optimal make-take fees for market making regulation

CHAPTER IX

Optimal make-take fees for market making regulation

Abstract

We consider an exchange who wishes to set suitable make-take fees to attract liquidity on its platform. Using a principal-agent approach, we are able to describe in quasi-explicit form the optimal contract to propose to a market maker. This contract depends essentially on the market maker inventory trajectory and on the volatility of the asset. We also provide the optimal quotes that should be displayed by the market maker. The simplicity of our formulas allows us to analyze in details the effects of optimal contracting with an exchange, compared to a situation without contract. We show in particular that it leads to higher quality liquidity and lower trading costs for investors.

Keywords: Make-take fees, market making, financial regulation, high-frequency trading, principal-agent problem, stochastic control.

1 Introduction

With the fragmentation of financial markets, exchanges are nowadays in competition. Indeed the traditional international exchanges are now challenged by alternative trading venues, see [LL13]. Consequently, they have to find innovative ways to attract liquidity on their platforms. One solution is to use a make-taker fees system, that is a rule enabling them to charge in an asymmetric way liquidity provision and liquidity consumption. The most classical setting, used by many exchanges (such as Nasdaq, Euronext, BATS Chi-X...), is of course to subsidize the former while taxing the latter. In practice, this means associating a fee rebate to executed limit orders and applying a transaction cost for market orders.

In the recent years, the topic of make-take fees has been quite controversial. Indeed make-take fees policies are seen as a major facilitating factor to the emergence of a new type of market makers aiming at collecting fee rebates: the high frequency traders. As stated by the Securities and Exchanges commission in [S.E10]: “Highly automated exchange systems and liquidity rebates have helped establish a business model for a new type of professional liquidity provider that is distinct from the more traditional exchange specialist and over-the-counter market maker." The concern with high frequency traders becoming the new liquidity providers is two-fold. First, their presence implies that slower traders no longer have access to the limit order book, or only in unfavorable situations when high frequency traders do not wish to support liquidity. This leads to the second classical criticism against high frequency market makers: they tend to leave the market in time of stress, see [Bel17, MSLR17, Men13, RP12] for detailed investigations about high frequency market making activity.

From an academic viewpoint, studies of make-take fees structures and their impact on the welfare of the markets have been mostly empirical, or carried out in rather stylized models. An interesting theory, suggested in [AHS11] and developed in [CF12] is that make-take fees have actually no impact on trading costs in the sense that the *cum fee* bid-ask spread should not depend on the make-take fees policy. This result is consistent with the empirical findings in [Lut10, MP15a]. Nevertheless, it is clearly shown in these works that many important trading parameters such as depths, volumes or price impact do depend on the make-take fees structure, see also [Har13]. Furthermore, the idea of the neutrality of the make-take fees schedule is also tempered in [FKK13] where the authors show theoretically that make-take fees may increase welfare of markets provided the tick size is not equal to zero, see also [BM13].

In this work, our aim is to provide a quantitative and operational answer to the question of relevant make-take fees. To do so, we take the position of an exchange (or of the regulator) wishing to attract liquidity. The exchange is looking for the best make-take fees policy to offer to market makers in order to maximize its utility. In other words, it aims at designing an optimal contract with the market marker to create an incentive to increase liquidity. For simplicity, we consider a single market maker in a non-fragmented market.

Incentive theory has emerged in the 1970s in economics to model how a financial agent can delegate the management of an output process to another agent. Let us recall the formalism of principal-agent problems from the seminal works of Mirrlees [Mir74] and Holmström [Höl79]. A principal aims at contracting with an agent who provides efforts to manage an output process impacting the wealth of the principal. The principal is not able to control directly the output process since he cannot decide the efforts made by the agent. In our case, the principal is the exchange, the agent is the market maker, the efforts correspond to the quality of the liquidity provided by the market maker (essentially the size of the bid-ask spread proposed by the market maker) and the output process is the transactions flow on the platform. Several economics papers have investigated this kind of problems by identifying it with a Stackelberg equilibrium between the two parties. More precisely, since the principal cannot control the work of the agent, he anticipates his best-reaction effort for a given compensation. Knowing that, the principal aims at finding the best contract.

In our work, we deal with a continuous-time principal-agent problem. Indeed, the exchange monitors the spread set by the market maker around a Brownian-type efficient price and the transactions flow in continuous-time. Thus, we follow the stream of literature initiated in

[HM87]. Then in [San08], the author recasts such issue into a stochastic control problem which has been further developed using backward stochastic differential equation theory in [CPT15]. See also [CZ13] for related literature.

In this paper, although we work in a quite general and realistic setting, we are able to solve our principal-agent problem. More precisely, we provide a quasi-explicit expression for the optimal contract the exchange should propose to the market maker, and also for the quotes the market maker should set. The optimal contract depends essentially on the market maker inventory trajectory and on the volatility of the market. These simple formulas enable us to analyze in details the effects for the welfare of the market of optimal contracting with an exchange, compared to a situation without contract as in [AS08, GLFT13]. We notably show that using such contracts leads to reduced spreads and lower trading costs for investors.

The paper is organized as follows. Our modeling approach is presented in Section 2. In particular, we define the market maker's as well as the exchange's optimization framework. In Section 3, we compute the best response of the market maker for a given contract. Optimal contracts are designed in Section 4 where we solve the exchange's problem. Then, in Section 5, we assess the benefits for market quality of the presence of an exchange contracting optimally with a market maker. Finally, useful technical results are gathered in an appendix.

2 The model

The framework considered throughout this paper is inspired by the seminal work [AS08] where the authors consider the problem of optimal market making, but without the intervention of an exchange. Let $T>0$ be a final horizon time and $(\Omega,\mathscr{F})$ be a measurable space such that $\Omega=\Omega_c\times(\Omega_d)^2$ with Ω_c the set of continuous functions from $[0,T]$ into $\mathbb{R}$, Ω_d the set of piecewise constant càdlàg functions from $[0,T]$ into $\mathbb{N}$ and $\mathscr{F}$ the Borel algebra on Ω. We consider the following canonical process $(\chi_t)_{t\in[0,T]}=(S_t,N_t^a,N_t^b)_{t\in[0,T]}$

$$\forall\omega=(s,n^a,n^b)\in\Omega\quad S_t(\omega)=s(t),\quad N_t^a(\omega)=n^a(t),\quad N_t^b(\omega)=n^b(t).$$

We endow the space $(\Omega,\mathscr{F})$ with $\mathbb{F}=(\mathscr{F}_t)_{t\in[0,T]}=(\mathscr{F}_t^c\otimes(\mathscr{F}_t^d)^{\otimes 2})_{t\in[0,T]}$ where $(\mathscr{F}_t^c)_{t\in[0,T]}$ and $(\mathscr{F}_t^d)_{t\in[0,T]}$ are the right-continuous completed filtrations associated with the components of $(\chi_t)_{t\in[0,T]}$.

We consider a market where there is only one market maker. This market maker has a view on the efficient price of the asset given by S_t. We assume that

$$S_t=S_0+\sigma W_t,\quad t\in[0,T],\tag{1}$$

with $S_0>0$, W a Brownian motion and $\sigma>0$ the volatility of the price[1]. For $t\in[0,T]$, the market maker fixes the bid and ask prices P_t^b and P_t^a as follows

$$P_t^b=S_t-\delta_t^b\quad\text{and}\quad P_t^a=S_t+\delta_t^a.$$

[1]In practice, the efficient price can be thought of as the mid-price of the asset.

We assume that the arrival of ask (resp. bid) market orders is modeled by a point process $(N_t^a)_{t\in[0,T]}$ (resp. $(N_t^b)_{t\in[0,T]}$) with intensity $(\lambda_t^a)_{t\in[0,T]}$ (resp. $(\lambda_t^b)_{t\in[0,T]}$). We also suppose that the volume of market orders is constant and equal to unity. Hence, the inventory process of the market maker Q is given by

$$Q_t = N_t^b - N_t^a, \quad t \in [0,T].$$

As in [GLFT13], we impose a critical absolute inventory $\bar{q} \in \mathbb{N}$ above which the market maker stops quoting on the ask or bid side, i.e.

$$\lambda_t^a = \lambda_t^a \mathbb{1}_{\{Q_t > -\bar{q}\}}, \quad \text{and} \quad \lambda_t^b = \lambda_t^b \mathbb{1}_{\{Q_t < \bar{q}\}}.$$

We expect the intensity of buy (resp. sell) market order arrivals to depend on the extra cost of each trade payed by the market taker compared to the efficient price. This extra cost is the sum of the spread δ_t^a (resp. δ_t^b) imposed by the market maker and the transaction cost $c > 0$ collected by the exchange, as explained in Section 2.2. Moreover, we recall that from classical financial economics results, see [DR15, MRR97, WBK$^+$08], the average number of trades per unit of time is a decreasing function of the ratio between the spread and the volatility. Hence, we assume that

$$\lambda_t^a = \lambda(\delta_t^a)\mathbb{1}_{\{Q_t > -\bar{q}\}}, \quad \text{and} \quad \lambda_t^b = \lambda(\delta_t^b)\mathbb{1}_{\{Q_t < \bar{q}\}}, \quad \text{with} \quad \lambda(x) = Ae^{-k\frac{(x+c)}{\sigma}}, \tag{2}$$

for fixed positive constants A and k.

2.1 Admissible controls and market maker's problem

We work with the set $\mathscr{A}$ of admissible controls $(\delta_t)_{t\in[0,T]} = (\delta_t^a, \delta_t^b)_{t\in[0,T]}$ where any $\delta \in \mathscr{A}$ is predictable and satisfies

$$|\delta_t^a| \vee |\delta_t^b| \quad \leq \quad \delta_\infty, \quad t \in [0,T].$$

Here, δ_∞ is a fixed positive constant which will be fixed later to a sufficiently large value. For each control process $\delta = (\delta^a, \delta^b)$ of the market maker, we denote by $\mathbb{P}^\delta$ the associated probability measure under which S follows (1) and

$$\widetilde{N}_t^{\delta,a} = N_t^a - \int_0^t \lambda(\delta_r^a)\mathbb{1}_{\{Q_r > -\bar{q}\}}dr, \quad \widetilde{N}_t^{\delta,b} = N_t^b - \int_0^t \lambda(\delta_r^b)\mathbb{1}_{\{Q_r < \bar{q}\}}dr,$$

are martingales. In that case, the profit and loss process of the market maker is defined by

$$\mathrm{PL}_t^\delta = X_t^\delta + Q_t S_t, \quad \text{where} \quad X_t^\delta = \int_0^t P_r^a dN_r^a - \int_0^t P_r^b dN_r^b, \quad t \in [0,T]. \tag{3}$$

Here, X^δ is the cash flow process and QS represents the inventory risk process[2].

[2] As in [AS08], for sake of simplicity, we assume that the market maker estimates his inventory risk using the efficient price S.

Next, we introduce the Doléans-Dade exponential

$$\begin{aligned} L_t^\delta &= \exp\Big(\int_0^t \log\Big(\frac{\lambda(\delta_r^a)}{A}\Big)\mathbb{I}_{\{Q_{r-}>-\bar{q}\}}dN_r^a + \log\Big(\frac{\lambda(\delta_r^b)}{A}\Big)\mathbb{I}_{\{Q_{r-}<\bar{q}\}}dN_r^b \\ &\qquad -(\lambda(\delta_r^a)-A)\mathbb{I}_{\{Q_r>-\bar{q}\}}dr - (\lambda(\delta_r^b)-A)\mathbb{I}_{\{Q_r<\bar{q}\}}dr\Big), \end{aligned}$$

which is a $\mathbb{P}^0$−local martingale[3] as it can be verified by direct application of Itô's formula that

$$dL_t^\delta = L_t^\delta\Big(\frac{\lambda(\delta_t^a)-A}{A}\mathbb{I}_{\{Q_{t-}>-\bar{q}\}}d\widetilde{N}_t^{0,a} + \frac{\lambda(\delta_t^b)-A}{A}\mathbb{I}_{\{Q_{t-}<\bar{q}\}}d\widetilde{N}_t^{0,b}\Big).$$

Since δ^a and δ^b are uniformly bounded, this local martingale satisfies the Novikov-type criterion in [Sok13] and thus is a martingale. From Theorem III.3.11 in [JS13], it follows that

$$\frac{d\mathbb{P}^\delta}{d\mathbb{P}^0}\Big|_{\mathcal{F}_t} = L_t^\delta, \quad \text{for all} \quad t \in [0,T]. \tag{4}$$

In particular, all the probability measures $\mathbb{P}^\delta$ indexed by $\delta \in \mathcal{A}$ are equivalent. We therefore use the notation $a.s$ for almost surely without ambiguity. We shall write $\mathbb{E}_t^\delta$ for the conditional expectation with respect to $\mathcal{F}_t$ with probability measure $\mathbb{P}^\delta$.

We consider that the exchange is compensated for each market order arrival and so aims at keeping the market liquid. Thus, we assume that it proposes to the market maker a contract, defined by an $\mathcal{F}_T$-measurable random variable ξ, in order to create an incentive to attract liquidity on the platform by reducing his spread. In addition to the realized profit and loss (3) on $[0,T]$, the market maker receives a compensation ξ from the exchange at the final time T, thus leading to the maximization problem,

$$\begin{aligned} V_{\mathrm{MM}}(\xi) = \sup_{\delta\in\mathcal{A}} J_{\mathrm{MM}}(\delta,\xi) \quad \text{where} \quad J_{\mathrm{MM}}(\delta,\xi) &= \mathbb{E}^\delta\Big[-e^{-\gamma(\xi+\mathrm{PL}_T^\delta-\mathrm{PL}_0^\delta)}\Big] \\ &= \mathbb{E}^\delta\Big[-e^{-\gamma\left(\xi+\int_0^T(\delta_t^a dN_t^a+\delta_t^b dN_t^b+Q_t dS_t)\right)}\Big]. \end{aligned} \tag{5}$$

Here, $\gamma > 0$ is the absolute risk aversion parameter of the CARA market maker. For each compensation ξ, we show that there exists a unique optimal response $\hat{\delta}(\xi) = (\hat{\delta}^a(\xi), \hat{\delta}^b(\xi))$ of the market marker.

Remark 1. *The case $\xi = 0$ corresponds to the problem without exchange intervention treated in [AS08, GLFT13].*

2.2 The exchange optimal contracting problem

We assume that the exchange is compensated by a fixed amount $c > 0$ for each market order that occurs in the market. In practice, some exchanges add to this fixed fee a component

[3] $\mathbb{P}^0$ denotes the the probability measure $\mathbb{P}^\delta$ associated to a vanishing spread $\delta = (\delta^a, \delta^b) = (0,0)$.

which is proportional to the traded amount in currency value. However, since we are anyway working on a short time interval, we take c independent of the price of the asset. Note that the fee schedule considered here for the taker side is simple. Indeed, in practice, complex fee policies are rather dedicated to market makers. Furthermore, we will in fact see that when acting optimally, the exchange is somehow indifferent to the value of c, see Section 4.3.

The exchange aims at maximizing the total number of market orders $N_T^a - N_0^a + N_T^b - N_0^b$ arrived during the time interval $[0,T]$, whose arrival intensities are controlled exclusively by the market maker. The role of the contract ξ proposed by the exchange to the market maker is to encourage the latter to increase the liquidity of the market. In this case, the profit and loss of the exchange is given by

$$c(N_T^a - N_0^a + N_T^b - N_0^b) - \xi.$$

Thus the exchange optimally chooses the contract to maximize its CARA utility function with absolute risk aversion parameter $\eta > 0$,

$$V_0^E \quad = \quad \sup_{\xi \in \mathscr{C}} \mathbb{E}^{\hat{\delta}(\xi)}\Big[- e^{-\eta(c(N_T^a - N_0^a + N_T^b - N_0^b) - \xi)}\Big]. \tag{6}$$

We now define the set of admissible contracts $\mathscr{C}$. Concerning the problem of the exchange, we need to ensure that $\mathbb{E}^{\hat{\delta}(\xi)}\Big[- e^{-\eta(c(N_T^a - N_0^a + N_T^b - N_0^b) - \xi)}\Big]$ is not degenerated. The natural condition that we need is then to assume that

$$\sup_{\delta \in \mathscr{A}} \mathbb{E}^{\delta}\Big[e^{\eta' \xi}\Big] < +\infty, \quad \text{for some} \quad \eta' > \eta. \tag{7}$$

Since N^a and N^b are point processes with bounded intensities, this condition together with a Hölder inequality ensure that the problem of the exchange (6) is well defined. Similarly, we will assume that

$$\sup_{\delta \in \mathscr{A}} \mathbb{E}^{\delta}\Big[e^{-\gamma' \xi}\Big] < +\infty, \quad \text{for some} \quad \gamma' > \gamma, \tag{8}$$

to ensure that $\mathbb{E}^{\delta}[-e^{-\gamma(\xi + \int_0^T (\delta_t^a dN_t^a + \delta_t^b dN_t^b + Q_t dS_t))}]$ is not degenerate and hence the well-definition of the market maker problem (5). We will also assume that the latter only accepts contracts ξ such that the maximal utility $V_{\mathrm{MM}}(\xi)$ is above a threshold value $R < 0$.

Hence, we denote by $\mathscr{C}$ the space of admissible contracts defined by

$$\mathscr{C} \quad = \quad \Big\{\xi \ \mathscr{F}_T\text{-measurable such that } V_{\mathrm{MM}}(\xi) \geq R \text{ and (7) and (8) are satisfied}\Big\}.$$

We will take $-R$ large enough so that $\mathscr{C}$ contains the zero contract $\xi = 0$ and thus is nonempty.

3 Solving the market maker's problem

We start by solving the problem (5) of the market maker facing an arbitrary contract $\xi \in \mathscr{C}$ proposed by the exchange.

For $(\delta, z, q) \in [-\delta_\infty, \delta_\infty]^2 \times \mathbb{R}^3 \times \mathbb{Z}$, with $\delta = (\delta^a, \delta^b)$ and $z = (z^S, z^a, z^b)$, we define

$$h(\delta, z, q) \quad = \quad \frac{1 - e^{-\gamma(z^a+\delta^a)}}{\gamma} \lambda(\delta^a) \mathbb{1}_{\{q > -\bar{q}\}} + \frac{1 - e^{-\gamma(z^b+\delta^b)}}{\gamma} \lambda(\delta^b) \mathbb{1}_{\{q < \bar{q}\}},$$

and

$$H(z, q) \quad = \quad \sup_{|\delta^a| \vee |\delta^b| \le \delta_\infty} h(\delta, z, q),$$

For an arbitrary constant $Y_0 \in \mathbb{R}$ and predictable processes $Z = (Z^S, Z^a, Z^b)$, with $\int_0^T |Z_t^S|^2 + |H(Z_t, Q_t)| dt < \infty$, we introduce the process

$$Y_t^{Y_0, Z} \;=\; Y_0 + \int_0^t Z_r^a dN_r^a + Z_r^b dN_r^b + Z_r^S dS_r + \Big(\frac{1}{2}\gamma\sigma^2(Z_r^S + Q_r)^2 - H(Z_r, Q_r)\Big) dr, \tag{9}$$

and we denote by $\mathcal{Z}$ the collection of all such processes Z such that Condition (7) is satisfied with $\xi = Y_T^{0,Z}$ and

$$\sup_{\delta \in \mathcal{A}} \sup_{t \in [0,T]} \mathbb{E}^\delta [e^{-\gamma' Y_t^{0,Z}}] < \infty, \quad \text{for some} \quad \gamma' > \gamma. \tag{10}$$

Clearly, $\mathcal{Z} \neq \emptyset$ as it contains all bounded predictable processes and

$$\mathcal{C} \;\supset\; \Xi = \big\{ Y_T^{Y_0, Z} : \; Y_0 \in \mathbb{R},\; Z \in \mathcal{Z}, \text{ and } V_{\mathrm{MM}}(Y_T^{Y_0, Z}) \ge R \big\}.$$

The next result shows that these sets are in fact equal, and identifies the market maker utility value and the corresponding optimal response. To prove equality of these sets, we are reduced to the problem of representing any contract $\xi \in \mathcal{C}$ as $\xi = Y_T^{Y_0, Z}$ for some $(Y_0, Z) \in \mathbb{R} \times \mathcal{Z}$, which is known in the literature as a problem of backward stochastic differential equation. We refrain from using this terminology, as our analysis does not require any result from this literature.

Theorem 1. *(i) Any contract $\xi \in \mathcal{C}$ has a unique representation as $\xi = Y_T^{Y_0, Z}$, for some $(Y_0, Z) \in \mathbb{R} \times \mathcal{Z}$. In particular, $\mathcal{C} = \Xi$.*
(ii) Under this representation, the market maker utility value is

$$V_{MM}\big(\xi\big) = -e^{-\gamma Y_0}, \quad \textit{so that} \quad \Xi = \Big\{ Y_T^{Y_0, Z} : \; Z \in \mathcal{Z}, \textit{ and } Y_0 \ge \frac{-1}{\gamma} \log(-R) \Big\},$$

with the following optimal bid-ask policy

$$\hat{\delta}_t^a(\xi) = \Delta(Z_t^a), \; \hat{\delta}_t^b(\xi) = \Delta(Z_t^b), \; \textit{where } \Delta(z) = (-\delta_\infty) \vee \Big\{ -z + \frac{1}{\gamma} \log\Big(1 + \frac{\sigma\gamma}{k}\Big) \Big\} \wedge \delta_\infty. \tag{11}$$

The proof of Part (i) of the previous result is reported in Section IX.B. This representation is obtained by using the dynamic continuation utility process of the market maker, following the approach of Sannikov [San08]. We prove that the continuation utility process satisfies the dynamic programming principle, so that the required representation follows from the Doob-Meyer decomposition of supermartingales together with the martingale representation theorem.

Proof of Theorem 1 (ii) Let $\xi = Y_T^{Y_0,Z}$ with $(Y_0, Z) \in \mathbb{R} \times \mathcal{Z}$. We first prove that for an arbitrary bid-ask policy $\delta \in \mathcal{A}$, we have $J_{\mathrm{MM}}(\delta, \xi) \leq -e^{-\gamma Y_0}$. Denote $\overline{Y}_t = Y_t^{Y_0,Z} + \int_0^t \delta_t^a dN_t^a + \delta_t^b dN_t^b + Q_t dS_t$, $t \in [0, T]$. By direct application of Itô's formula, we see that

$$de^{-\gamma \overline{Y}_t} = \gamma e^{-\gamma \overline{Y}_{t-}} \Big(-(Q_t + Z_t^S) dS_t - \frac{1}{\gamma}(1 - e^{-\gamma(Z_t^a + \delta_t^a)}) d\widetilde{N}_t^{\delta,a} - \frac{1}{\gamma}(1 - e^{-\gamma(Z_t^b + \delta_t^a)}) d\widetilde{N}_t^{\delta,b}$$
$$+ \big(H(Z_t, Q_t) - h(\delta_t, Z_t, Q_t)\big) dt \Big).$$

Hence $e^{-\gamma \overline{Y}}$ is a $\mathbb{P}^\delta$-local submartingale. Thanks to Condition (10), the uniform boundedness of the intensities of N^a and N^b and Hölder inequality, $(e^{-\gamma \overline{Y}_t})_{t \in [0,T]}$ is uniformly integrable and hence is a true submartingale. By Doob-Meyer decomposition theorem, we conclude that

$$\int_0^\cdot \gamma e^{-\gamma \overline{Y}_{t-}} \big(-(Q_t + Z_t^S) dS_t - \frac{1}{\gamma}(1 - e^{-\gamma(Z_t^a + \delta_t^a)}) d\widetilde{N}_t^{\delta,a} - \frac{1}{\gamma}(1 - e^{-\gamma(Z_t^b + \delta_t^a)}) d\widetilde{N}_t^{\delta,b} \big),$$

is a true martingale. It follows that

$$J_{\mathrm{MM}}(\delta, \xi) = \mathbb{E}^\delta\big[-e^{-\gamma \overline{Y}_T} \big] = -e^{-\gamma Y_0} - \mathbb{E}^\delta \Big[\int_0^T \gamma e^{-\gamma \overline{Y}_t} \big(H(Z_t, Q_t) - h(\delta_t, Z_t, Q_t) \big) dt \Big] \leq -e^{-\gamma Y_0}.$$

On the other hand, equality holds in the last inequality if and only if δ is chosen as the maximizer of the Hamiltonian H ($dt \times d\mathbb{P}^0$−a.e.), thus leading to the unique maximizer $\hat{\delta}(\xi)$ given by (11), which then induces $J_{\mathrm{MM}}(\hat{\delta}(\xi), \xi) = -e^{-\gamma Y_0}$. This completes the proof that $V_{\mathrm{MM}}(\xi) = -e^{-\gamma Y_0}$ with optimal response $\hat{\delta}(\xi)$. □

4 Designing the optimal contract

Denote $\hat{Y}_0 = -\frac{1}{\gamma}\log(-R)$. By Theorem 1, the exchange problem (6) reduces to the control problem

$$V_0^E = \sup_{Y_0 \geq \hat{Y}_0} \sup_{Z \in \mathcal{Z}} \mathbb{E}^{\hat{\delta}(Y_T^{Y_0,Z})} \Big[-e^{-\eta\big(c(N_T^a - N_0^a + N_T^b - N_0^b) - Y_T^{Y_0,Z}\big)} \Big], \tag{12}$$

where $Y^{Y_0,Z}$ is given by (9). In the present context, notice that the market maker optimal response $\hat{\delta}(Y_T^{Y_0,Z})$ given by (11) does not depend on Y_0, *i.e* $\hat{\delta}(Y_T^{Y_0,Z}) = \hat{\delta}(Y_T^{\hat{Y}_0,Z})$. Hence, the objective function in (12) is clearly decreasing in Y_0 implying that the maximization under the participation constraint is achieved at $\hat{Y}_0$,

$$V_0^E = e^{\eta \hat{Y}_0} \sup_{Z \in \mathcal{Z}} \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})} \Big[-e^{-\eta\big(c(N_T^a - N_0^a + N_T^b - N_0^b) - Y_T^{0,Z}\big)} \Big]. \tag{13}$$

4.1 The HJB equation for the reduced exchange problem

Motivated by (13), we study in this section the HJB equation corresponding to the stochastic control problem

$$v_0^E \quad = \quad \sup_{Z \in \mathcal{Z}} \ \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})} \Big[-e^{-\eta\big(c(N_T^a - N_0^a + N_T^b - N_0^b) - Y_T^{0,Z}\big)} \Big]. \tag{14}$$

Our approach is to derive a solution v of the corresponding HJB equation, and to proceed by the standard verification argument in stochastic control to prove that the proposed solution v coincides with the value function v_0^E.

Applying the standard dynamic programming approach to the last control problem, we are led to the following HJB equation

$$\begin{cases} \partial_t v(t,q) + H_E\big(q, v(t,q), v(t,q+1), v(t,q-1)\big) \;=\; 0, \quad q \in \{-\bar{q}, \cdots \bar{q}\}, \quad t \in [0,T), \\ v(T,q) = -1, \end{cases} \tag{15}$$

where the Hamiltonian $H_E : [-\bar{q}, \bar{q}] \times (-\infty, 0]^3 \to \mathbb{R}$ is given by

$$H_E(q, y, y_+, y_-) \;=\; H_E^1(q,y) + \mathbb{1}_{\{q > -\bar{q}\}} H_E^0(y, y_-) + \mathbb{1}_{\{q < \bar{q}\}} H_E^0(y, y_+), \tag{16}$$

with

$$H_E^1(q,y) \;=\; \sup_{z_s \in \mathbb{R}} h_E^1(q, y, z_s), \;\text{ and }\; h_E^1(q,y,z_s) = \frac{\eta\sigma^2}{2}\, y\big(\gamma(z_s+q)^2 + \eta z_s^2\big),$$

$$H_E^0(y,y') \;=\; \sup_{\zeta \in \mathbb{R}} h_E^0(y,y',\zeta) \;\text{ and }\; h_E^0(y,y',\zeta) = \lambda\big(\Delta(\zeta)\big)\Big[y' e^{\eta(\zeta-c)} - y\big(1+\eta\,\frac{1-e^{-\gamma(\zeta+\Delta(\zeta))}}{\gamma}\big)\Big].$$

A direct calculation reported in Lemma 4 below reveals that the maximizers $\hat{z} = (\hat{z}^s, \hat{z}^a, \hat{z}^s)$ of H_E are

$$\hat{z}^s(t,q) = -\frac{\gamma}{\gamma+\eta} q, \;\; \hat{z}^a(t,q) = \hat{\zeta}\big(v(t,q), v(t,q-1)\big), \text{ and } \hat{z}^b(t,q) = \hat{\zeta}\big(v(t,q), v(t,q+1)\big), \tag{17}$$

where

$$\hat{\zeta}(y,y') \;=\; \zeta_0 + \frac{1}{\eta}\log\Big(\frac{y}{y'}\Big), \;\; \zeta_0 \;=\; c + \frac{1}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big).$$

Here, we assume that δ_∞ is large enough so that Condition (37) of Lemma 4 is always met, namely

$$\delta_\infty \geq C_\infty + \frac{1}{\eta} \sup_{t \in [0,T]} \sup_{q \in [-\bar{q}, \bar{q}-1]} \left| \log\Big(\frac{v(t,q)}{v(t,q+1)}\Big) \right| \tag{18}$$

with the hope that our candidate solution of the HJB equation will verify it. This will be checked in our verification argument. Recall from Lemma 4 that

$$C_\infty = c + (\frac{1}{\eta} + \frac{1}{\gamma}) \log(1 + \frac{\sigma\gamma}{k}) - \frac{1}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big).$$

Using again the calculation reported in Lemma 4, we rewrite the HJB equation (15) as

$$\begin{cases} \partial_t v(t,q) + \frac{\gamma\eta^2\sigma^2}{2(\gamma+\eta)} q^2 v(t,q) - C v(t,q)\Big[\mathbb{1}_{\{q>-\bar{q}\}}\big(\frac{v(t,q)}{v(t,q-1)}\big)^{\frac{k}{\sigma\eta}} + \mathbb{1}_{\{q<\bar{q}\}}\big(\frac{v(t,q)}{v(t,q+1)}\big)^{\frac{k}{\sigma\eta}}\Big] \;=\; 0, \\ v(T,q) = -1, \end{cases} \tag{19}$$

where the constant C is given by

$$C \quad = \quad A\frac{\sigma\eta}{k}\exp\Big(-\frac{k}{\sigma\gamma}\log(1+\frac{\sigma\gamma}{k})+(1+\frac{k}{\sigma\eta})\log\big(1-\frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\big)\Big).$$

Inspired by [GLFT13], we now make the key observation that this equation can be reduced to a linear equation by introducing $u=(-v)^{-\frac{k}{\sigma\eta}}$. Indeed, by direct substitution, we obtain the following linear differential equation

$$\begin{cases}\partial_t u(t,q)-C_1q^2u(t,q)+C_2\big(u(t,q+1)\mathbb{1}_{\{q<\bar{q}\}}+u(t,q-1)\mathbb{1}_{\{q>-\bar{q}\}}\big) \;=\; 0, \quad t\in[0,T),\\ u(T,q)=1,\end{cases} \tag{20}$$

with

$$C_1 \;=\; \frac{k\gamma\eta\sigma}{2(\gamma+\eta)} \quad \text{and} \quad C_2=C\frac{k}{\sigma\eta}.$$

This equation can be written in terms of the $\mathbb{R}^{2\bar{q}+1}$–valued function $\mathbf{u}(t)=\big(u(t,q)\big)_{q\in\{-\bar{q},\dots,\bar{q}\}}$, of the variable t only, as the linear ordinary differential equation

$$\partial_t\mathbf{u}=-\mathbf{B}u, \quad \text{where} \quad \mathbf{B}=\begin{pmatrix} -C_1\bar{q}^2 & C_2 & & & \\ \ddots & \ddots & \ddots & & \\ & C_2 & -C_1q^2 & C_2 & \\ & & \ddots & \ddots & \ddots \\ & & & C_2 & -C_1\bar{q}^2\end{pmatrix} \begin{array}{l} \\ \\ \leftarrow q\text{-th line,} \\ \\ \\ \end{array}$$

is a tri-diagonal matrix with lines labelled $-\bar{q},\dots,\bar{q}$. Denote by $\mathbf{b}_q$ the vector of $\mathbb{R}^{2\bar{q}+1}$ with zeros everywhere except at the position q, i.e. $\mathbf{b}_{q,i}=\mathbb{1}_{\{i=q\}}$ for $i\in\{-\bar{q},\dots,\bar{q}\}$, and $\mathbf{1}=\sum_{q=-\bar{q}}^{\bar{q}}\mathbf{b}_q$. Then, this ODE has a unique solution

$$\mathbf{u}(t)=e^{(T-t)\mathbf{B}}\mathbf{1}, \text{ so that } u(t,q)=\mathbf{b}_q\cdot e^{(T-t)\mathbf{B}}\mathbf{1}, \text{ and } v(t,q)=-\big(\mathbf{b}_q\cdot e^{(T-t)\mathbf{B}}\mathbf{1}\big)^{-\frac{\sigma\eta}{k}}. \tag{21}$$

In the next section, we shall prove that this solution v of the HJB equation (15) coincides with the value function of the reduced exchange problem (14), with optimal controls $\hat{z}(t,q)$ given in (17), thus inducing the optimal contract $Y_T^{\hat{Y}_0,\hat{Z}}$ with $\hat{Z}_t=\hat{z}(t,Q_{t-})$.

We conclude this section by an alternative representation of the function u.

Proposition 1. *Let u and v be defined by* (21). *The function u can be represented as*

$$u(t,q) \quad = \quad \mathbb{E}\Big[e^{\int_t^T(-C_1(Q_s^{t,q})^2+\overline{\lambda}_s+\underline{\lambda}_s)ds}\Big],$$

where $Q_s^{t,q}=q+\int_t^s d(\overline{N}_u-\underline{N}_u)$, and $(\overline{N},\underline{N})$ is a two-dimensional point process with intensity $(\overline{\lambda}_s,\underline{\lambda}_s)=C_2(\mathbb{1}_{\{Q_{s-}<\bar{q}\}},\mathbb{1}_{\{Q_{s-}>-\bar{q}\}})$. In particular, we have the following bounds for the function u,

$$e^{-C_1\bar{q}^2T}\le u\le e^{2C_2T}.$$

Moreover, Condition (18) *is verified when*

$$\delta_\infty \;\geq\; \Delta_\infty = C_\infty + \frac{\sigma}{k}(2C_2 + C_1\bar{q}^2)T. \tag{22}$$

Proof. Notice that u is a smooth bounded function. Denote $f(x) = -C_1x^2 + C_2(\mathbb{1}_{\{x>-\bar{q}\}} + \mathbb{1}_{\{x<\bar{q}\}})$, and $M_s = e^{\int_t^s f(Q_u^{t,q})du}u(s,Q_s^{t,q})$, $t \leq s \leq T$. We now show that M is a martingale, so that $u(t,q) = M_t = \mathbb{E}[M_T] = \mathbb{E}\big[e^{-\int_t^T f(Q_s^{t,q})ds}\big]$, as $u(T,.) = 1$. To see that M is a martingale, we compute by Itô's formula that

$$\begin{aligned} dM_s &= \big[u(s,Q_s^{t,q})f(Q_s^{t,q}) + \partial_t u(s,Q_s^{t,q})\big]ds \\ &\quad + C_2\big[u(s,Q_{s-}^{t,q}+1) - u(s,Q_{s-}^{t,q})\big]d\overline{N}_s + C_2\big[u(s,Q_{s-}^{t,q}-1) - u(s,Q_{s-}^{t,q})\big]d\underline{N}_s. \end{aligned}$$

Since u is solution of (20), we get

$$dM_s \;=\; C_2\big[u(s,Q_{s-}^{t,q}+1) - u(s,Q_{s-}^{t,q})\big]d\overline{M}_s + C_2\big[u(s,Q_{s-}^{t,q}-1) - u(s,Q_{s-}^{t,q})\big]d\underline{M}_s,$$

where $(\overline{M},\underline{M}) = (\overline{N} - \int_0^{\cdot}\overline{\lambda}_s ds, \underline{N} - \int_0^{\cdot}\underline{\lambda}_s ds)$ is a martingale. The martingale property of M now follows from the boundedness of u as it can be verified from the expression (21). Finally, the bound $|Q_s^{t,q}| \leq \bar{q}$ induces directly the announced bounds on u, which in turn imply Condition (18) when (22) is satisfied because $v = -u^{-\frac{\sigma\eta}{k}}$. □

4.2 Main result

We are now ready to verify that the function v introduced in the previous section is the value function of the exchange, with optimal feedback controls $(\hat{z}^s, \hat{z}^a, \hat{z}^b)$ as given in (17), thus identifying a unique optimal contract to be proposed by the exchange to the market maker. Recall that δ_∞ denotes the bound on the market maker bid and ask spreads. Our main explicit solution requires δ_∞ to be larger that the constant Δ_∞ introduced in (22).

Theorem 2. *Assume that $\delta_\infty \geq \Delta_\infty$, with Δ_∞ given by* (22) *and define u and v by* (21)*. Then the optimal contract for the problem of the exchange* (6) *is given by*

$$\hat{\xi} \;=\; \hat{Y}_0 + \int_0^T \hat{Z}_r^a dN_r^a + \hat{Z}_r^b dN_r^b + \hat{Z}_r^S dS_r + \Big(\frac{1}{2}\gamma\sigma^2\big(\hat{Z}_r^S + Q_r\big)^2 - H\big(\hat{Z}_r, Q_r\big)\Big)dr, \tag{23}$$

with $\hat{Z}_r^S = \hat{z}^s(r,Q_{r-})$, $\hat{Z}_r^a = \hat{z}^a(r,Q_{r-})$, and $\hat{Z}_r^b = \hat{z}^b(r,Q_{r-})$ as defined in (17)*. The market maker's optimal effort is given by*

$$\hat{\delta}_t^a = \hat{\delta}_t^a(\hat{\xi}) = -\hat{Z}_t^a + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}), \quad \hat{\delta}_t^b = \hat{\delta}_t^b(\hat{\xi}) = -\hat{Z}_t^b + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}). \tag{24}$$

Proof. In order to prove this result, we verify that the function v introduced in (21) coincides at $(0,Q_0)$ with the value function of the reduced exchange problem (14), with maximum achieved at the optimal control $\hat{Z}$.

The function v is negative bounded and has bounded gradient. Moreover, since $\delta_\infty \geq \Delta_\infty$, it follows that v is a solution of the HJB equation (15) of the exchange reduced problem, see Lemma 4. For $Z \in \mathcal{Z}$, denote

$$K_t^Z \;=\; e^{-\eta\left(c(N_t^a-N_0^a+N_t^b-N_0^b)-Y_t^{0,Z}\right)}, \quad t\in[0,T].$$

By direct application of Itô's formula, and substitution of $\partial_t v$ from the HJB equation satisfied by v, we see that

$$\begin{aligned} d\big[v(t,Q_t)K_t^Z\big] = K_{t-}^Z\Big(&(h_t^Z-\mathcal{H}_t)dt+\eta v(t,Q_t)Z_t^s dS_t \\ &+\sum_{i=a,b}\big[v(t,Q_{t-}+\Delta Q_t)e^{-\eta(c-Z_t^i)}-v(t,Q_{t-})\big]d\widetilde{N}_t^{\hat{\delta}(Y^{\hat{Y}_0,Z}),i}\Big), \end{aligned} \tag{25}$$

where, using the notations of (16) and the subsequent equations,

$$\mathcal{H}_t \;=\; H_E\big(Q_t,v(t,Q_t),v(t,Q_t+1),v(t,Q_t-1)\big),$$

and

$$\begin{aligned} h_t^Z = {}& h_E^1\big(Q_t,v(t,Q_t),Z_t^S\big)+\mathbb{1}_{\{Q_t>-\bar{q}\}}h_E^0\big(v(t,Q_t),v(t,Q_t-1),Z_t^a\big) \\ &+\mathbb{1}_{\{Q_t<\bar{q}\}}h_E^0\big(v(t,Q_t),v(t,Q_t+1),Z_t^b\big). \end{aligned}$$

Exploiting the fact that v is bounded and that K^Z is uniformly integrable, see Lemma 5, we get that $\big(v(t,Q_t)K_t^Z\big)_{t\in[0,T]}$ is a $\mathbb{P}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}$-supermartingale and by Doob-Meyer decomposition theorem, the local martingale term in (25) is a true martingale. Hence

$$\begin{aligned} v(0,Q_0) \;&=\; \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}\Big[v(T,Q_T)K_T^Z+\int_0^T K_t^Z(\mathcal{H}_t-h_t^Z)dt\Big] \\ &\geq\; \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}\big[v(T,Q_T)K_T^Z\big] \;=\; \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[-K_T^Z], \end{aligned}$$

by the boundary condition $v(T,.)=-1$. By arbitrariness of $Z\in\mathcal{Z}$, this provides the inequality $v(0,Q_0)\geq\sup_{Z\in\mathcal{Z}}\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}0,Z})}[-K_T^Z]=v_0^E$.

On the other hand, consider the maximizer $\hat{Z}$ of the reduced exchange problem, induced by the feedback controls $\hat{z}$ in (17). As $\hat{Z}$ is bounded, it follows that $\hat{Z}\in\mathcal{Z}$. Moreover, $h^{\hat{Z}}-\mathcal{H}=0$, by definition, so that the last argument leads to the equality $v(0,Q_0)=\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,\hat{Z}})}\big[-K_T^{\hat{Z}}\big]$, instead of the inequality. This shows that $v(0,Q_0)=v_0^E$, the reduced exchange problem of (14), with optimal control $\hat{Z}$. From Theorem 1, the corresponding optimal market maker response of the market maker is given by (11) with $\xi=Y_T^{\hat{Y}_0,\hat{Z}}$. Moreover, Condition (18) implies that

$$\left|-Z_t^i+\frac{1}{\gamma}\log\left(1+\frac{\sigma k}{k}\right)\right|\leq\delta_\infty, \quad i=a,b.$$

Hence the optimal effort could be reduced to (24). □

4.3 Discussion

The processes $\hat{Z}^a$, $\hat{Z}^b$ and $\hat{Z}^S$ allowing the exchange to build the optimal contract have actually quite natural interpretations. Indeed, using Lemma 1, we obtain that the quantities

$$-\log\Big(\frac{u(t,Q_{t-})}{u(t,Q_{t-}-1)}\Big) \text{ and } -\log\Big(\frac{u(t,Q_{t-})}{u(t,Q_{t-}+1)}\Big)$$

are roughly proportional respectively to Q_{t-} and $-Q_{t-}$. Thus, when the inventory is highly positive, the exchange provides incentives to the market-maker so that it attracts buy market orders and tries to dissuade him to accept more sell market orders, and conversely for a negative inventory. The integral

$$\int_0^T \hat{Z}_r^S dS_r$$

can be understood as a risk sharing term. Indeed, $\int_0^t Q_r dS_r$ corresponds to the price driven component of the inventory risk $Q_t S_t$. Hence in the optimal contract, the exchange supports part of this risk so that the market maker maintains reasonable quotes despite some inventory. The proportion of risk handled by the platform is $\frac{\gamma}{\gamma+\eta}$.

Until now, we have focused on the maker part of the make-take fees problem since we have considered that the taker cost c is fixed. Nevertheless, our approach also enables us to suggest the exchange a relevant value for c. Actually, we see that when acting optimally, the exchange transfers the totality of the fixed taker fee c to the market maker. It is therefore neutral to the value of c as its optimal utility function $v_0^E = v(0,Q_0)$ is independent of the taker cost, see (19). However, c plays an important role in the optimal spread offered by the market maker given by

$$-2c+\frac{\sigma}{k}\log\Big(\frac{u(t,Q_{t-})^2}{u(t,Q_{t-}-1)u(t,Q_{t-}+1)}\Big)-\frac{2}{\eta}\log\Big(1-\frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big)+\frac{2}{\gamma}\log(1+\frac{\sigma\gamma}{k}).$$

Furthermore, from numerical computations, we remark that

$$\frac{u(t,q)^2}{u(t,q-1)u(t,q+1)}$$

is close to unity for any t and q. Hence the exchange may fix in practice the transaction cost c so that the spread is close to one tick by setting

$$c \approx -\frac{1}{2}\text{Tick}-\frac{1}{\eta}\log\Big(1-\frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big)+\frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}).$$

For $\sigma\gamma/k$ small enough, this equation reduces to

$$c \approx \frac{\sigma}{k}-\frac{1}{2}\text{Tick}. \tag{26}$$

Equation (26) is a particularly simple formula to fix the taker constant c. We see that the higher the volatility, the larger the taker cost should be. It is also quite natural that this cost is a decreasing function of k. Indeed, if k is large, the liquidity vanishes rapidly when the spread becomes wide, meaning that market takers are sensitive to extra costs relative to the efficient price. Therefore, the taker cost has to be small if the exchange wants to maintain a reasonable market order flow. Finally, note that the parameters σ and k can be easily estimated from market data. Therefore the formula (26) can be readily used in practice.

5 Impact of the presence of the exchange on market quality and comparison with [AS08, GLFT13]

In this section, we compare our setting with the situation without incentive policy from an exchange towards market making activities. The latter is considered in [AS08, GLFT13] where the authors deal with the issue of optimal market making without intervention of the exchange. The results in [AS08] are taken as benchmark for our investigation to emphasize the impact of the incentive policy on market quality. We will refer to this case as the neutral exchange case.

Let us first recall the results in [AS08, GLFT13]. The optimal controls of the market maker denoted by $\widetilde{\delta}^a$ and $\widetilde{\delta}^b$ are given as a function of the inventory Q_t by

$$\widetilde{\delta}^a_t = \frac{\sigma}{k}\log\Big(\frac{\widetilde{u}(t,Q_{t-})}{\widetilde{u}(t,Q_{t-}-1)}\Big) + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}),$$
$$\widetilde{\delta}^b_t = \frac{\sigma}{k}\log\Big(\frac{\widetilde{u}(t,Q_{t-})}{\widetilde{u}(t,Q_{t-}+1)}\Big) + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}),$$

where $\widetilde{u}$ is the unique solution of the linear differential equation

$$\begin{cases}\partial_t\widetilde{u}(t,q)+\widetilde{C_1}q^2\widetilde{u}(t,q)-\widetilde{C_2}(\widetilde{u}(t,q+1)\mathbb{1}_{\{q<\bar{q}\}}+\widetilde{u}(t,q-1)\mathbb{1}_{\{q>-\bar{q}\}})=0, (t,q)\in[0,T)\times[-\bar{q},\bar{q}]\\ \widetilde{u}(T,q)=1,\end{cases}$$

with $\widetilde{C_1}=\frac{\sigma\gamma k}{2}$ and $\widetilde{C_2}=A\exp\big(-(1+\frac{\sigma\gamma}{k})\log(1+\frac{\sigma\gamma}{k})\big)$. In our case, the optimal quotes $\hat{\delta}^a$ and $\hat{\delta}^b$ are obtained from Theorem 2 and satisfy

$$\hat{\delta}^a_t = \frac{\sigma}{k}\log\Big(\frac{u(t,Q_{t-})}{u(t,Q_{t-}-1)}\Big) + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}) - c - \frac{1}{\eta}\log\Big(1-\frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big),$$
$$\hat{\delta}^b_t = \frac{\sigma}{k}\log\Big(\frac{u(t,Q_{t-})}{u(t,Q_{t-}+1)}\Big) + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}) - c - \frac{1}{\eta}\log\Big(1-\frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big),$$

where u is solution of the linear equation (20).

Numerical experiments show that u and $\widetilde{u}$ can decrease quickly to zero when q becomes large. Hence, the computation of the following crucial quantities appearing in the optimal quotes:

$$v_+(t,q)=\log\Big(\frac{u(t,q+1)}{u(t,q)}\Big),\quad \widetilde{v}_+(t,q)=\log\Big(\frac{\widetilde{u}(t,q+1)}{\widetilde{u}(t,q)}\Big),\quad q\in\{-\bar{q},\cdots,\bar{q}-1\}.$$

can be intricate in practice. To circumvent this numerical difficulty, we remark that v_+ and $\widetilde{v}_+$ are solution of the following differential equations

$$\begin{cases} \partial_t v_+(t,q) + C_1(2q+1) - C_2(e^{v_+(t,q+1)}\mathbb{1}_{\{q<\bar{q}-1\}} + e^{-v_+(t,q)} - e^{v_+(t,q)} - e^{-v_+(t,q-1)}\mathbb{1}_{\{q>-\bar{q}\}}) = 0 \\ v_+(T,q) = 0, \end{cases} \tag{27}$$

and

$$\begin{cases} \partial_t \widetilde{v}_+(t,q) + \widetilde{C_1}(2q+1) - \widetilde{C_2}(e^{\widetilde{v}_+(t,q+1)}\mathbb{1}_{\{q<\bar{q}-1\}} + e^{-\widetilde{v}_+(t,q)} - e^{\widetilde{v}_+(t,q)} - e^{-\widetilde{v}_+(t,q-1)}\mathbb{1}_{\{q>-\bar{q}\}}) = 0 \\ \widetilde{v}_+(T,q) = 0. \end{cases} \tag{28}$$

We thus rather apply classical finite difference schemes to (27) and (28).

In the following numerical illustrations, in the spirit of [GLFT13, Section 6], we take $T = 600s$ for an asset with volatility $\sigma = 0.3$ Tick.$s^{-1/2}$ (unless specified differently). Market orders arrive according to the intensities (2) with $A = 1.5s^{-1}$ and $k = 0.3s^{-1/2}$. We assume that the threshold inventory of the market maker is $\bar{q} = 50$ units and we set his risk aversion parameter to $\gamma = 0.01$. The exchange is taken more risk averse with $\eta = 1$. Finally, we assume that the taker cost $c = 0.5$ Tick[4].

5.1 Impact of the exchange on the spread and market liquidity

We start by comparing the optimal spread $\hat{\delta}_0^a + \hat{\delta}_0^b$ at time 0 obtained when contracting optimally with the spread without incentives towards market making activities $\widetilde{\delta}_0^a + \widetilde{\delta}_0^b$. The optimal spreads are plotted in Figure IX.1 for different initial inventory values $Q_0 \in \{-\bar{q}, \cdots, \bar{q}\}$.

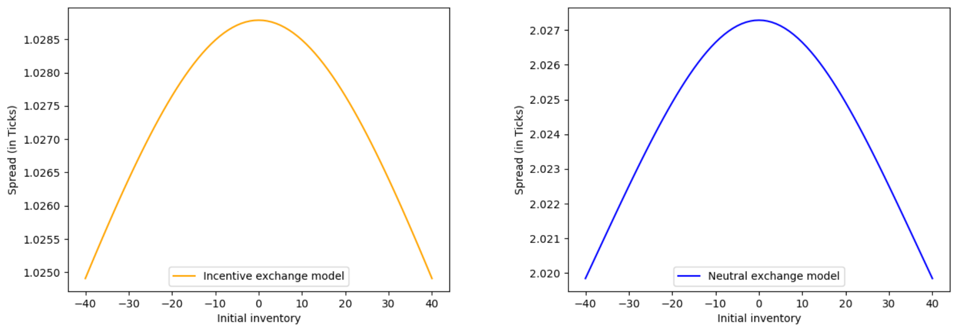

Figure IX.1 – Comparison of optimal initial spreads with/without incentive policy from the exchange.

We observe in Figure IX.1 that the initial spread does not depend a lot on the initial inventory (because the considered time interval $[0, T]$ is not too small) and that it is reduced thanks to

[4]Note that the taker cost is chosen according to Criteria (26). We expect the optimal spread to be close to one tick.

the optimal contract between the market maker and the exchange. This is not surprising since in our case the exchange aims at increasing the market order flow by proposing an incentive contract to the market maker inducing a spread reduction. Actually this phenomenon occurs over the whole trading period $[0, T]$. To see this, we generate 5000 paths of market scenarios and compute the average spread over $[0, T]$ for an initial inventory $Q_0 = 0$. The results are given in Figure IX.2.

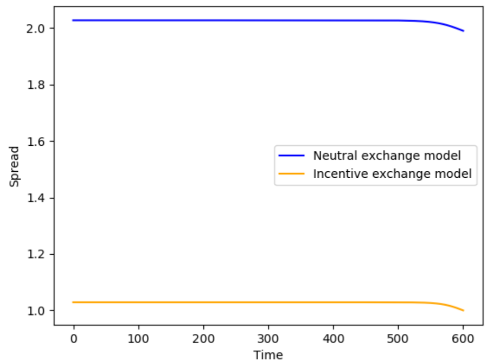

Figure IX.2 – Average spread on $[0, T]$ with 95% confidence interval, with/without incentive policy from the exchange toward the market maker.

Since the spread is tighter during the trading period under an incentive policy from the exchange, the arrival intensity of market orders is more important and hence the market is more liquid as shown in Figure IX.3.

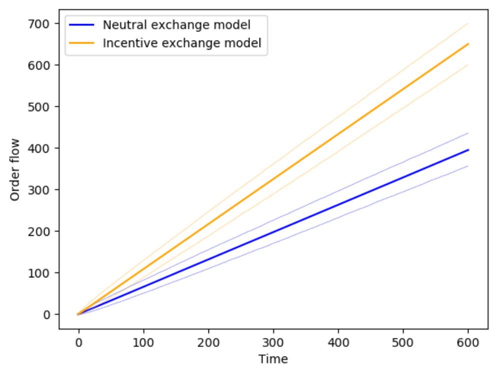

Figure IX.3 – Average order flow on $[0, T]$ with 95% confidence interval, with/without incentive policy from the exchange.

We now consider in Figure IX.4 the bid and ask sides separately. We see that when the inventory is positive and very large, $\hat{\delta}^a$ and $\widetilde{\delta}^a$ are negative. It means the market maker is ready to sell at prices lower than the efficient price in order to attract market orders and reduce his inventory risk. On the contrary, if the inventory is negative and very large, in both situations, its ask quotes are well above the efficient price in order to repulse the arrival of buy market orders. However, since in our case the exchange remunerates the market maker for each arrival of market order, we get that the ask spread with contract $\hat{\delta}^a$ is smaller than $\widetilde{\delta}^a$. A symmetric conclusion holds for the bid part of the spread.

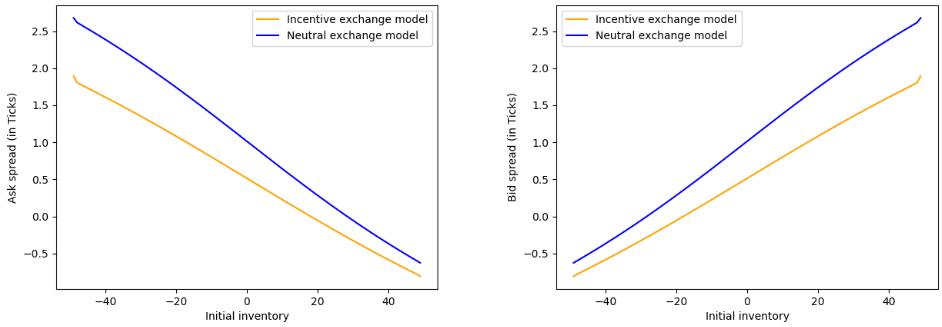

Figure IX.4 – Optimal ask and bid spreads, with/without incentive policy from the exchange toward the market maker.

We now turn to the impact of the volatility on the spread. The optimal contract obtained in (23) induces an inventory risk sharing phenomenon through the term $\hat{Z}^S$. Hence, when the volatility increases, the spread difference between situations with/without incentive policy becomes less important, see Figure IX.5 in which we consider the optimal initial spread difference when the initial inventory is set to zero between both situations with/without incentive policy from the exchange to the market maker for different values of the volatility.

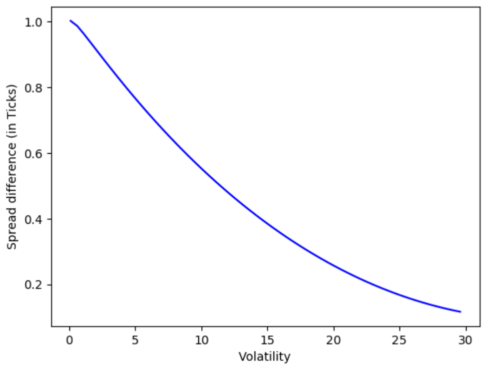

Figure IX.5 – The initial optimal spread difference between both situation with/without incentive policy from the exchange toward the market maker as a decreasing function of the volatility σ.

5.2 Impact of the incentive policy on the profit and loss of the exchange and market maker

We assume that $Q_0 = 0$. Recall that PL^{δ} defined in (3) denotes the trading part of the profit and loss (P&L) of the market maker for a given strategy δ. In our case, the underlying total P&L at time t of a market maker acting optimally, denoted by $\mathrm{PL}^{\star}_t$, is given by

$$\mathrm{PL}^{\star}_t = \mathrm{PL}^{\hat{\delta}}_t + Y_t^{\hat{Y}_0,\hat{Z}},$$

where $Y_t^{\hat{Y}_0,\hat{Z}}$ corresponds to the quantity on the right hand side of (23) with T replaced by t. We now investigate the behavior of this quantity, notably with respect to the benchmark $\mathrm{PL}^{\tilde{\delta}}_t$ which corresponds to the optimal profit and loss without intervention of the exchange.

To make $\mathrm{PL}^{\star}_t$ and $\mathrm{PL}^{\tilde{\delta}}_t$ comparable, we choose $\hat{Y}_0$ in (23) so that the market maker gets the same utility in both situations, that is $\hat{Y}_0 = \frac{k}{\sigma}\log(\tilde{u}(0,Q_0))$. Thus, the market maker is indifferent between the situation with or without exchange intervention. We generate 5000 paths of market scenarios and compare the average of both P&L in Figure IX.6 with and without incentive policy.

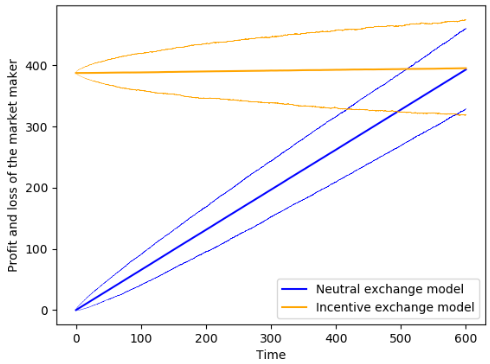

Figure IX.6 – Average P&L of the market maker on $[0,T]$ with 95% confidence interval, with/without incentive policy from the exchange.

Since $\hat{Y}_0$ is set to obtain the same utility in both cases, the two average P&L are very close at the end of the trading period. The variance of the P&L also seems to be the same in both situations. The only difference from the market maker viewpoint here is that in the case of a contract, the P&L is already made at time 0 thanks to the compensation of the exchange and then fluctuates slightly. This is because he is earning the spread but paying continuous "coupons" $\big(H(\hat{Z}_t,Q_t) - \frac{\sigma^2\gamma}{2}(\hat{Z}^S_t + Q_t)^2\big)dt$ from the contract. In the case without exchange intervention, the market maker increases his P&L over the whole trading period thanks to the spread.

We now compare the profit and loss of the exchange in the two considered cases. When it applies an incentive policy towards the market maker, the P&L of the exchange is given by

$c(N_t^a - N_0^a + N_t^b - N_0^b) - Y_t^{\hat{Y}_0, \hat{Z}}$. When the exchange is neutral, its P&L is simply $c(N_t^a - N_0^a + N_t^b - N_0^b)$. We compare these two quantities in Figure IX.7.

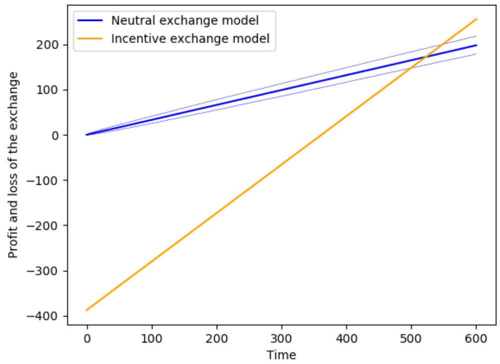

Figure IX.7 – Average P&L of the exchange on $[0, T]$ with 95% confidence interval, with/without incentive policy from the exchange.

We see that the initial P&L of the contracting exchange is negative because of the initial payment $\hat{Y}_0$. However it finally exceeds, with a smaller standard deviation, the P&L in the situation without incentive policy from the exchange. Hence the incentive policy of the exchange proves to be successful. Indeed, both configurations are equivalent for market makers but the exchange obtains more revenues when contracting optimally. This is due to the fact that the contract triggers more market orders.

Finally, we plot the aggregated average P&L of the market maker and the exchange (independent of the choice of the initial payment). We observe that it is always greater in the optimal contract case.

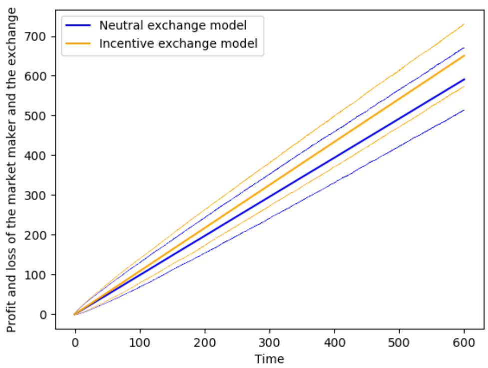

Figure IX.8 – Average total P&L of the exchange and the market maker on $[0, T]$ with 95% confidence interval, with/without incentive policy from the exchange.

5.3 Impact of the incentive policy on the trading cost

We now study the impact of the incentive policy on the investors, viewed as the market takers. We assume that there is only one market taker. In the case without exchange, with the specified parameters and under optimal reaction of the market maker, this investor buys on average 200 shares over $[0, T]$. To make the comparison with the case with exchange intervention, we modify the parameter A appearing in the intensity (2) when simulating a market with optimal contract. This new value is chosen so that the investor buys on average the same number of assets (200) over the time period. This amounts to take $A = 0.9s^{-1}$. We confirm in Figure IX.9 that the average ask order flows agree in both situations.

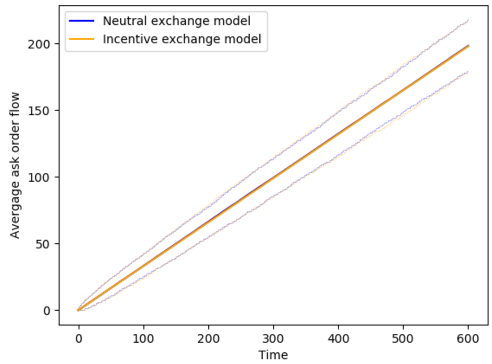

Figure IX.9 – Setting similar average ask order flows on $[0, T]$ by taking different intensity basis A in the case with and in the case without incentive policy; 95% confidence interval.

We finally compare in Figure IX.10 the average cost of trading for the market taker

$$\mathbb{E}^{\delta}\Big[\int_0^T \delta_t^a dN_t^a\Big],$$

with and without incentive.

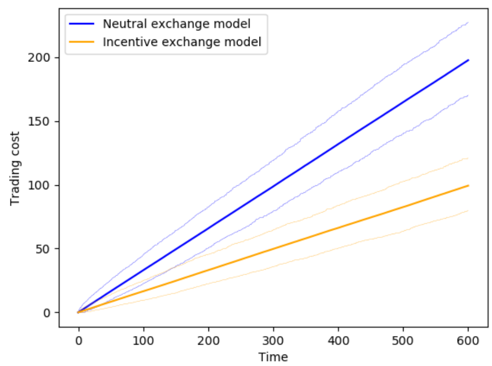

Figure IX.10 – Average trading cost on $[0,T]$ with 95% confidence interval, with/without incentive policy from the exchange.

We see that, thanks to the incentive policy of the exchange, the reduced spreads lead to significantly smaller trading costs for investors.

IX.A Predictable representation

The following result is probably well-known, we report it here for completeness as we could not find a precise reference.

Lemma 1. *Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a filtered probability space where $\mathbb{F} = \mathbb{F}^W \vee \mathbb{F}^N$ is the right continuous completed filtration of a Brownian motion W and a d-dimensional integrable point process $N = (N^1, \cdots, N^d)$ with compensator $A = (A^1, \cdots, A^d)$. Then, for any $\mathbb{F}-$martingale X there exists a predictable process $Z = (Z^W, Z^1, \cdots, Z^d)$ such that*

$$X_t = X_0 + \int_0^t Z_s^W dW_s + \sum_{i=1}^d \int_0^t Z_s^i (dN_s^i - dA_s^i).$$

Proof. For sake of simplicity, we take $d = 1$. Let $\mathbb{P}$ be a solution of the martingale problem associated to $M_t = N_t - A_t$ and W_t. By Theorem III.4.29 in [JS13], to prove Lemma 1 we need to establish the uniqueness of $\mathbb{P}$.

We denote by $\mathbb{P}^W$ the law $\mathbb{P}$ conditional on W. We first show that M is still a martingale under $\mathbb{P}^W$. To do so we consider $B_s \in \mathcal{F}_s$ and want to prove that

$$\mathbb{E}^{\mathbb{P}^W}\left[\mathbb{1}_{B_s}(M_t - M_s)\right] = 0,$$

for $0 \le s \le t \le T$. Let $C \in \mathcal{F}_T^W$. We aim at showing that

$$\mathbb{E}\left[1_C \mathbb{E}^{\mathbb{P}^W}\left[\mathbb{1}_{B_s}(M_t - M_s)\right]\right] = \mathbb{E}\left[\mathbb{1}_C \mathbb{1}_{B_s}(M_t - M_s)\right] = 0.$$

Thanks to the martingale representation theorem for Brownian martingales, we can write

$$\mathbb{1}_C = \alpha_s + \int_s^T \phi_u dW_u,$$

where $\alpha_s = \mathbb{E}[\mathbb{1}_C | \mathcal{F}_s^W]$ and $(\phi_u)_{u \ge 0}$ is $\mathbb{F}^W$ predictable process. Using the martingale property of M, we obtain

$$\mathbb{E}\left[\alpha_s \mathbb{1}_{B_s}(M_t - M_s)\right] = 0.$$

Then W and M being orthogonal martingales, we deduce

$$\mathbb{E}\left[\int_s^T \phi_u dW_u \mathbb{1}_{B_s}(M_t - M_s)\right] = 0,$$

Consequently, using Theorem III.1.21 in [JS13], $\mathbb{P}^W$ is the unique probability measure such that M is an $\mathbb{F}$-martingale conditional on W. Finally, by integration, the uniqueness of $\mathbb{P}^W$ implies that of $\mathbb{P}$. □

IX.B Dynamic programming principle and contract representation

IX.B.1 Dynamic programming principle

For all $\mathbb{F}$-stopping time τ with values in $[t,T]$ and for any $\mu \in \mathscr{A}_\tau$, we define[5]

$$J_T(\tau,\mu) = \mathbb{E}_\tau^\mu\left[-e^{-\gamma\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)}e^{-\gamma\xi}\right], \quad \text{and} \quad \mathscr{J}_{\tau,T} = \big(J_T(\tau,\mu)\big)_{\mu\in\mathscr{A}_\tau},$$

where $\mathscr{A}_\tau$ denotes the restriction of $\mathscr{A}$ to controls on $[\tau,T]$. The continuation utility of the market maker is defined for all $\mathbb{F}$-stopping time τ by

$$V_\tau \quad = \quad \operatorname*{ess\,sup}_{\mu\in\mathscr{A}_\tau} J_T(\tau,\mu).$$

Lemma 2. *Let τ be an $\mathbb{F}$-stopping time with values in $[t,T]$. Then, there exists a non-decreasing sequence $(\mu^n)_{n\in\mathbb{N}}$ in $\mathscr{A}_\tau$ such that $V_\tau = \lim_{n\to+\infty}\uparrow J_T(\tau,\mu^n)$.*

Proof. For μ and μ' in $\mathscr{A}_\tau$, define $\hat\mu = \mu\mathbb{1}_{\{J_T(\tau,\mu)\geq J_T(\tau,\mu')\}} + \mu'\mathbb{1}_{\{J_T(\tau,\mu)<J_T(\tau,\mu')\}}$. Then $\hat\mu \in \mathscr{A}_\tau$ and by definition of $\hat\mu$

$$J_T(\tau,\hat\mu) \geq \max\big(J_T(\tau,\mu), J_T(\tau,\mu')\big).$$

Hence $\mathscr{J}_{\tau,T}$ is directly upwards, and the required result folows from [Nev72, Proposition VI.I.I p121]. □

Lemma 3. *Let $t \in [0,T]$ and τ be an $\mathbb{F}$-stopping time with values in $[t,T]$. Then,*

$$V_t \quad = \quad \operatorname*{ess sup}_{\delta\in\mathscr{A}} \mathbb{E}_t^\delta\Big[-e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}V_\tau\Big].$$

Proof. Let $t\in[0,T]$ and set an $\mathbb{F}$-stopping time τ with values in $[t,T]$. The proof is similar to [CK93, Proof of Proposition 6.2]. First, by the tower property,

$$V_t \quad = \quad \operatorname*{ess\,sup}_{\delta\in\mathscr{A}} \mathbb{E}_t^\delta\left[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}\mathbb{E}_\tau^\delta\left[-e^{-\gamma\left(\int_\tau^T(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)+\xi\right)}\right]\right].$$

For all $\delta\in\mathscr{A}$, the quotient $\frac{L_T^\delta}{L_\tau^\delta}$ does not depend on the values of δ before time τ. Then,

$$\begin{aligned}\mathbb{E}_\tau^\delta\left[-e^{-\gamma\left(\int_\tau^T(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)+\xi\right)}\right] \quad &= \quad \mathbb{E}_\tau^0\left[-\frac{L_T^\delta}{L_\tau^\delta}e^{-\gamma\left(\int_\tau^T(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)+\xi\right)}\right]\\ &\leq \quad \operatorname*{ess\,sup}_{\mu\in\mathscr{A}_\tau}\mathbb{E}_\tau^\mu\left[-e^{-\gamma\left(\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)+\xi\right)}\right] \; = \; V_\tau,\end{aligned}$$

[5]From (4), notice that for any $\delta\in\mathscr{A}$, the conditional expectation $\mathbb{E}_\tau^\delta$ depends only on the restriction of δ on $[\tau,T]$. Hence $\mathbb{E}_\tau^\mu$ is defined without ambiguity for $\mu\in\mathscr{A}_\tau$.

Then,

$$V_t \leq \operatorname*{ess\,sup}_{\delta\in\mathscr{A}} \mathbb{E}_t^\delta\Big[V_\tau e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}\Big].$$

We next prove the reverse inequality. Let $\delta\in\mathscr{A}$ and $\mu\in\mathscr{A}_\tau$. We define $(\delta\otimes_\tau\mu)_u=\delta_u\mathbf{1}_{0\leq u<\tau}+\mu_u\mathbf{1}_{\tau\leq u\leq T}$. Then $\delta\otimes_\tau\mu\in\mathscr{A}$ and

$$\begin{aligned}V_t &\geq \mathbb{E}_t^{\delta\otimes_\tau\mu}\Big[-e^{-\gamma\left(\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)+\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)\right)}e^{-\gamma\xi}\Big]\\ &= \mathbb{E}_t^{\delta\otimes_\tau\mu}\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}\mathbb{E}_\tau^{\delta\otimes_\tau\mu}\Big[-e^{-\gamma\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)}e^{-\gamma\xi}\Big]\Big].\end{aligned} \tag{29}$$

From Bayes' Formula and by noticing that $\frac{L_T^{\delta\otimes_\tau\mu}}{L_\tau^{\delta\otimes_\tau\mu}}=\frac{L_T^\mu}{L_\tau^\mu}$, we get

$$\begin{aligned}\mathbb{E}_\tau^{\delta\otimes_\tau\mu}\Big[-e^{-\gamma\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)}e^{-\gamma\xi}\Big] &= \mathbb{E}_\tau^0\left[\frac{L_T^{\delta\otimes_\tau\mu}}{L_\tau^{\delta\otimes_\tau\mu}}\Big(-e^{-\gamma\int_\tau^T(\mu_u^a dN_u^a+\mu_u^b dN_u^b+Q_u dS_u)}e^{-\gamma\xi}\Big)\right]\\ &= J_T(\tau,\mu).\end{aligned}$$

Thus, Inequality (29) becomes

$$V_t \geq \mathbb{E}_t^{\delta\otimes_\tau\mu}\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\Big].$$

By using again Bayes' Formula and by noticing that $\frac{L_\tau^{\delta\otimes_\tau\mu}}{L_t^{\delta\otimes_\tau\mu}}=\frac{L_\tau^\delta}{L_t^\delta}$, we have

$$\begin{aligned}V_t &\geq \frac{\mathbb{E}_t^0\Big[L_T^{\delta\otimes_\tau\mu}e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\Big]}{L_t^{\delta\otimes_\tau\mu}}\\ &= \mathbb{E}_t^0\left[\mathbb{E}_\tau^0\Big[\frac{L_T^{\delta\otimes_\tau\mu}}{L_\tau^{\delta\otimes_\tau\mu}}\frac{L_\tau^{\delta\otimes_\tau\mu}}{L_t^{\delta\otimes_\tau\mu}}e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\Big]\right]\\ &= \mathbb{E}_t^0\left[\mathbb{E}_\tau^0\Big[\frac{L_T^{\delta\otimes_\tau\mu}}{L_\tau^{\delta\otimes_\tau\mu}}\Big]\frac{L_\tau^{\delta\otimes_\tau\mu}}{L_t^{\delta\otimes_\tau\mu}}e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\right]\\ &= \mathbb{E}_t^0\left[\frac{L_\tau^{\delta\otimes_\tau\mu}}{L_t^{\delta\otimes_\tau\mu}}e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\right]\\ &= \mathbb{E}_t^\delta\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu)\Big].\end{aligned}$$

Since the previous inequality holds for all $\mu\in\mathscr{A}_\tau$ we deduce from monotone convergence Theorem together with Lemma 2 that there exists a sequence $(\mu^n)_{n\in\mathbb{N}}$ of control in $\mathscr{A}_\tau$ such that

$$\begin{aligned}V_t &\geq \lim_{n\to+\infty}\uparrow\mathbb{E}_t^\delta\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}J_T(\tau,\mu^n)\Big]\\ &= \mathbb{E}_t^\delta\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}\lim_{n\to+\infty}\uparrow J_T(\tau,\mu^n)\Big] = \mathbb{E}_t^\delta\Big[e^{-\gamma\int_t^\tau(\delta_u^a dN_u^a+\delta_u^b dN_u^b+Q_u dS_u)}V_\tau\Big],\end{aligned}$$

thus concluding the proof. $\square$

IX.B.2 Proof of Theorem 1 (i)

We proceed in several steps.
Step 1. For $\delta \in \mathcal{A}$, it follows from the dynamic programming principle of Lemma 3 that the process

$$U_t^\delta = V_t e^{-\gamma \int_0^t (\delta_u^a dN_u^a + \delta_u^b dN_u^b + Q_u dS_u)}, \quad t \in [0,T],$$

defines a $\mathbb{P}^\delta$-supermartingale[6] for all $\delta \in \mathcal{A}$. By standard analysis, we may then consider it in its càdlàg version (by taking right limits along rationals). By the Doob-Meyer decomposition, we write

$$U_t^\delta = M_t^\delta - A_t^{\delta,c} - A_t^{\delta,d}, \tag{30}$$

where M^δ is a $\mathbb{P}^\delta$-martingale and $A^\delta = A^{\delta,c} + A^{\delta,d}$ is an integrable non-decreasing predictable process such that $A_0^{\delta,c} = A_0^{\delta,d} = 0$, with pathwise continuous component $A^{\delta,c}$, and a piecewise constant predictable process $A^{\delta,d}$.

By the martingale representation theorem under $\mathbb{P}^\delta$, see Lemma 1, there exists a predictable process $\widetilde{Z}^\delta = (\widetilde{Z}^{\delta,S}, \widetilde{Z}^{\delta,a}, \widetilde{Z}^{\delta,b})$ such that

$$M_t^\delta = V_0 + \int_0^t \widetilde{Z}_r^\delta . d\chi_r - \int_0^t \widetilde{Z}_r^{\delta,a} \lambda(\delta_r^a) \mathbb{1}_{\{Q_r > -\bar{q}\}} dr - \int_0^t \widetilde{Z}_r^{\delta,b} \lambda(\delta_r^b) \mathbb{1}_{\{Q_r < \bar{q}\}} dr, \tag{31}$$

where we recall that $\chi = (S, N^a, N^b)$.

Step 2. We show that V is a negative process. In fact, thanks to the uniform boundedness of $\delta \in \mathcal{A}$, we show that

$$\frac{L_T^\delta}{L_t^\delta} \geq \alpha_{t,T} = e^{-\frac{k\delta_\infty}{\sigma}(N_T^a - N_t^a + N_T^b - N_t^b) - 2Ae^{-\frac{kc}{\sigma}}(e^{\frac{k\delta_\infty}{\sigma}} + 1)(T-t)} > 0. \tag{32}$$

Therefore,

$$V_t \leq \mathbb{E}^0 \Big[-\alpha_{t,T} e^{-\gamma\left(\delta_\infty (N_T^a - N_t^a + N_T^b - N_t^b) + \int_t^T Q_u dS_u\right)} e^{-\gamma \xi} \Big] < 0.$$

Step 3. Let Y be the process defined by $V_t = -e^{-\gamma Y_t}$ for all $t \in [0,T]$. As $A^{\delta,d}$ is a predictable point process and the jumps of (N^a, N^b) are totally inaccessible stopping times under $\mathbb{P}^0$, we have $[N^a, A^{\delta,d}] = 0$ and $[N^b, A^{\delta,d}] = 0$ a.s, see Proposition I.2.24 in [JS13]. Using Itô's formula, we obtain from (30) and (31) that

$$Y_T = \xi, \quad \text{and} \quad dY_t = Z_t^a dN_t^a + Z_t^b dN_t^b + Z_t^S dS_t - dI_t - d\widetilde{A}_t^d,$$

[6] Note that $\mathbb{E}^\delta[U_T^\delta] = J_T(0,\delta) > -\infty$ using Hölder inequality together with (8) and the uniform boundedness of the intensities of N^a and N^b. Hence the process U^δ is integrable.

where $Z^a, Z^b, Z^S, I, \widetilde{A}^d$ are independent of δ, as they may be expressed as $Z^i_t dN^i_t = d[Y, N^i]_t$, $i \in \{a, b\}$, $Z^S_t \sigma^2 dt = d\langle Y_t, S_t\rangle_t$, $\widetilde{A}^d$ the predictable pure jumps of Y. Moreover, Itô's Formula yields

$$Z^a_t = -\frac{1}{\gamma}\log(1 + \frac{\widetilde{Z}^{\delta,a}_t}{U^\delta_{t-}}) - \delta^a_t, \quad Z^b_t = -\frac{1}{\gamma}\log(1 + \frac{\widetilde{Z}^{\delta,b}_t}{U^\delta_{t-}}) - \delta^b_t, \quad Z^S_t = -\frac{\widetilde{Z}^{\delta,b}_t}{\gamma U^\delta_{t-}} - Q_{t-},$$

and

$$I_t = \int_0^t \Big(\overline{h}(\delta_r, Z_r, Q_r)dr - \frac{1}{\gamma U^\delta_r} dA^{\delta,c}_r\Big), \qquad \widetilde{A}^d_t = \frac{1}{\gamma}\sum_{s\le t}\log\Big(1 - \frac{\Delta A^{\delta,d}_t}{U^\delta_{t-}}\Big),$$

with $\overline{h}(\delta, z, q) = h(\delta, z, q) - \frac{1}{2}\gamma\sigma^2(z^s)^2$. In particular, the last relation between $\widetilde{A}^d$ and $A^{\delta,d}$ shows that $\Delta a_t = \frac{-\Delta A^{\delta,d}_t}{U^\delta_{t-}} \ge 0$ is independent of $\delta \in \mathscr{A}$; recall that $U^\delta < 0$.

In order to complete the proof, we argue in the subsequent steps that $Z = (Z^S, Z^a, Z^b) \in \mathcal{Z}$ and that for $t \in [0, T]$,

$$A^{\delta,d}_t = -\sum_{s\le t} U^\delta_{s-}\Delta a_s = 0, \ \text{(so that } \widetilde{A}^d_t = 0\text{)}, \quad \text{and} \quad I_t = \int_0^t \overline{H}(Z_r, Q_r)dr, \tag{33}$$

where $\overline{H}(z, q) = H(z, q) - \frac{1}{2}\gamma\sigma^2(z^s)^2$.

Step 4. Since $V_T = -1$, notice that

$$\begin{aligned} 0 = \sup_{\delta\in\mathscr{A}} \mathbb{E}^\delta[U^\delta_T] - V_0 &= \sup_{\delta\in\mathscr{A}} \mathbb{E}^\delta[U^\delta_T - M^\delta_T] \\ &= \gamma \sup_{\delta\in\mathscr{A}} \mathbb{E}^0\Big[L^\delta_T \int_0^T U^\delta_{r-}\big(dI_r - \overline{h}(\delta_r, Z_r, Q_r)dr + \frac{da_r}{\gamma}\big)\Big]. \end{aligned} \tag{34}$$

Moreover, since the controls are uniformly bounded, we have

$$U^\delta_t \le -\beta_t := V_t e^{-\gamma\delta_\infty(N^a_t - N^a_0 + N^b_t - N^b_0) - \gamma\int_0^t Q_r dS_r} < 0. \tag{35}$$

Then, since $A^{\delta,d} \ge 0$, $U^\delta \le 0$, and $dI_t - \overline{h}(\delta_t, Z_t, Q_t) \ge 0$, it follows from (34) together with the inequalities (32) and (35) that

$$\begin{aligned} 0 &\le \sup_{\delta\in\mathscr{A}} \mathbb{E}^0\Big[\alpha_{0,T}\int_0^T -\beta_{r-}\big(dI_r - \overline{h}(\delta_r, Z_r, Q_r)dr + \frac{da_r}{\gamma}\big)\Big] \\ &= -\mathbb{E}^0\Big[\alpha_{0,T}\int_0^T \beta_{r-}\big(dI_r - \overline{H}(Z_r, Q_r)dr + \frac{da_r}{\gamma}\big)\Big]. \end{aligned}$$

The quantities $\alpha_{0,T}\int_0^T \beta_{r-}\big(dI_r - \overline{H}(Z_r, Q_r)dr\big)$ and $\alpha_{0,T}\int_0^T \beta_r da_r$ being non-negative random variables, this implies (33).

Step 5. We now prove that $Z \in \mathcal{Z}$ by showing that

$$\sup_{\delta\in\mathcal{A}} \sup_{t\in[0,T]} \mathbb{E}^\delta[e^{-\gamma(p+1)Y_t}] < \infty \quad \text{for some } p > 0. \tag{36}$$

Using Hölder inequality together with Condition (8) and the boundedness of the intensities of N^a and N^b, we have that $\sup_{\delta\in\mathcal{A}} \mathbb{E}^\delta[|U_T^\delta|^{p'+1}] < \infty$ for some $p' > 0$. Hence

$$\sup_{\delta\in\mathcal{A}} \sup_{t\in[0,T]} \mathbb{E}^\delta[|U_t^\delta|^{p'+1}] = \sup_{\delta\in\mathcal{A}} \mathbb{E}^\delta[|U_T^\delta|^{p'+1}] < \infty,$$

because U^δ is a negative P^δ-supermartingale. This leads to (36) using Hölder inequality, the uniform boundedness of the intensities of N^a and N^b and that $e^{-\gamma Y} = U^\delta e^{\gamma\int_0^\cdot(\delta_u^a dN_u^a + \delta_u^b dN_u^b + Q_u dS_u)}$.

Step 6. We finally prove uniqueness of the representation. Let $(Y_0, Z), (Y_0', Z') \in \mathbb{R}\times\mathcal{Z}$ be such that $\xi = Y_T^{Y_0,Z} = Y_T^{Y_0',Z'}$. By following the line of the verification argument in the proof of Theorem 1 (ii), we obtain the equality $Y_t^{Y_0,Z} = Y_t^{Y_0',Z'}$ by considering the value of the continuation utility of the market maker

$$-\exp(-\gamma Y_t^{Y_0,Z}) = -\exp(-\gamma Y_t^{Y_0',Z'}) = \operatorname*{ess\,sup}_{\delta\in\mathcal{A}} \mathbb{E}_t^\delta[-e^{-\gamma(\mathrm{PL}_T^\delta - \mathrm{PL}_t^\delta + \xi)}], \quad t\in[0,T].$$

This in turn implies that $Z_t^i dN_t^i = Z_t'^i dN_t^i = d[Y^{Y_0,Z}, N^i]_t$, $i\in\{a,b\}$, and $Z_t^S\sigma^2 dt = Z_t'^S\sigma^2 dt = d\langle Y, S\rangle_t$, $t\in[0,T]$. Hence $(Y_0, Z) = (Y_0', Z')$.

IX.C Exchange Hamiltonian maximization

Lemma 4. *Let $c\in\mathbb{R}$, $\gamma,\eta,k,\sigma > 0$ and $v_1, v_2 < 0$. We define for $z\in\mathbb{R}$*

$$\varphi(z) = Ae^{-k\frac{\Delta(z)+c}{\sigma}}\Big(v_1 e^{\eta(z-c)} - v_2\big(\frac{\eta}{\gamma}\big(1 - e^{-\gamma(z+\Delta(z))}\big) + 1\big)\Big),$$

with $\Delta(z) = (-\delta_\infty)\vee\big(-z + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})\big)\wedge\delta_\infty$ and $\delta_\infty > 0$. Provided

$$\delta_\infty \geq C_\infty + \frac{1}{\eta}|\log(\frac{v_2}{v_1})|, \tag{37}$$

with $C_\infty = |c| + (\frac{1}{\eta}+\frac{1}{\gamma})\log(1+\frac{\sigma\gamma}{k}) - \frac{1}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big)$, the function φ is nondecreasing on $(-\infty, -\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})]$ and non-increasing on $[\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}), \infty)$. It admits a maximum on $[-\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}), \delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})]$ attained in $z^\star$ given by

$$z^\star = c + \frac{1}{\eta}\log(v_2/v_1) + \frac{1}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big).$$

In that case, we have

$$\varphi(z^\star) = -Cv_2\exp\big(\frac{k}{\sigma\eta}\log(v_2/v_1)\big),$$

where

$$C = A\frac{\sigma\eta}{k}\exp\Big(-\frac{k}{\sigma\gamma}\log(1+\frac{\sigma\gamma}{k}) + (1+\frac{k}{\sigma\eta})\log\big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\big)\Big).$$

Proof. Easy but tedious computations lead to prove that φ is non-decreasing on $(-\infty, -\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})]$ and non-increasing on $[\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}), \infty)$ if,

$$\delta_\infty \geq \Big| c + \frac{1}{\eta}\log(v_2/v_1) - (\frac{1}{\eta}+\frac{1}{\gamma})\log(1+\frac{\sigma\gamma}{k})\Big|.$$

Moreover, we notice that φ admits a maximum on $[-\delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k}), \delta_\infty + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})]$ attained in

$$z^\star = c + \frac{1}{\eta}\log(v_2/v_1) + \frac{1}{\eta}\log\Big(1 - \frac{\sigma^2\gamma\eta}{(k+\sigma\gamma)(k+\sigma\eta)}\Big),$$

as soon as $\delta_\infty \geq |-z^\star + \frac{1}{\gamma}\log(1+\frac{\sigma\gamma}{k})|$. By combining these two conditions, we get the result under Condition (37) on δ_∞. □

IX.D On the verification argument for the exchange problem

The proof of the main result of Theorem 2 requires the following technical result. We observe that this is the place where Condition (7) is needed.

Lemma 5. *Let $Z \in \mathcal{Z}$. There exists $C > 0$ and $\varepsilon > 0$ such that*

$$\sup_{t\in[0,T]} \mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^{1+\varepsilon}] \leq C.$$

Proof. We recall the definition of K^Z for $Z \in \mathcal{Z}$

$$K_t^Z = e^{-\eta\left(c(N_t^a - N_0^a + N_t^b - N_0^b) - Y_t^{0,Z}\right)},\ t \in [0,T].$$

Let $p > 1$. By using Hölder's inequality and the uniform boundedness of the intensities of N^a and N^b, we deduce that there exists $C' > 0$ such that

$$\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^p] \leq C'\mathbb{E}^0[(e^{-\gamma Y_t^{0,Z}})^{-\frac{p'\eta}{\gamma}}]^{\frac{p}{p'}},$$

with any $p' > p$. Thus,

$$\begin{aligned}
\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^p] &\leq C'\left(1 + \mathbb{E}^0[(e^{-\gamma Y_t^{0,Z}})^{-\frac{p'\eta}{\gamma}}]\right)\\
&= C'\left(1 + \mathbb{E}^0\left[(-\sup_{\delta\in\mathcal{A}}\mathbb{E}_t^\delta[-e^{-\gamma(Y_T^{0,Z}+PL_T^\delta - PL_t^\delta)}])^{-\frac{p'\eta}{\gamma}}\right]\right).
\end{aligned}$$

From Jensen's inequality and then Hölder's inequality, we deduce that for any $p'' > p'$ we have

$$\begin{aligned}
\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^p] &\leq C'\left(1 + \mathbb{E}^0\left[\sup_{\delta\in\mathcal{A}}\mathbb{E}_t^\delta[e^{p'\eta(Y_T^{0,Z}+PL_T^\delta - PL_t^\delta)}]\right]\right)\\
&\leq C'\left(1 + \mathbb{E}^0\left[\sup_{\delta\in\mathcal{A}}\mathbb{E}_t^\delta[e^{p''\eta Y_T^{0,Z}}]\right]\right).
\end{aligned}$$

By using a dynamic programming principle, similarly to the proof of Lemma 3 by noticing that the family $\left(\widetilde{J}(\mu,\delta) = \mathbb{E}^{\delta}_{\tau}[e^{p''\eta Y^{0,Z}_T}]\right)_{\mu\in\mathscr{A}_\tau}$ is directly upwards, we get

$$\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^p] \leq C'\left(1+\sup_{\delta\in\mathscr{A}}\mathbb{E}^{\delta}\left[e^{p''\eta Y_T^{0,Z}}\right]\right).$$

By setting $\varepsilon = \frac{\eta'-\eta}{3}$, if we take $p = 1+\varepsilon$, then $p' = p+\varepsilon$ and $p'' = p'+\varepsilon$, we obtain

$$\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^{1+\varepsilon}] \leq C'\left(1+\sup_{\delta\in\mathscr{A}}\mathbb{E}^{\delta}\left[e^{\eta' Y_T^{0,Z}}\right]\right).$$

From the definition of $\mathcal{Z}$ (involving the condition (7)), we get for any $t\in[0,T]$

$$\mathbb{E}^{\hat{\delta}(Y_T^{\hat{Y}_0,Z})}[|K_t^Z|^{1+\varepsilon}] \leq C,$$

with $C = C'\left(1+\sup_{\delta\in\mathscr{A}}\mathbb{E}^{\delta}\left[e^{\eta' Y_T^{0,Z}}\right]\right) < +\infty$. $\square$

Bibliography

[AC01] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[AGR17] E. Alòs, J. Gatheral, and R. Radoičić. Exponentiation of conditional expectations under stochastic volatility. *SSRN*, 2017.

[AHS11] J. J. Angel, L. E. Harris, and C. S. Spatt. Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53, 2011.

[AJLP17] E. Abi Jaber, M. Larsson, and S. Pulido. Affine Volterra processes. *arXiv preprint arXiv:1708.08796*, 2017.

[ALR14] F. Abergel, C.-A. Lehalle, and M. Rosenbaum. Understanding the stakes of high-frequency trading. *The Journal of Trading*, 9(4):49–73, 2014.

[ALV07] E. Alòs, J. A. León, and J. 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.

[AMST07] H. Albrecher, P. Mayer, W. Schoutens, and J. Tistaert. The little Heston trap. *Wilmott Magazine*, pages 83–92, January 2007.

[AP07] L. B. Andersen and V. V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

[AS] E. Alos and K. Shiraya. Estimating the hurst parameter from short term volatility swaps.

[AS08] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

[ASCDL15] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.

[ASFL13] Y. Ait-Sahalia, J. Fan, and Y. Li. The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109(1):224–249, 2013.

[BBMZ05] L. Borland, J.-P. Bouchaud, J.-F. Muzy, and G. Zumbach. The dynamics of financial markets-Mandelbrot's cascades and beyond. *Wilmott Magazine*, 2005.

[BCK92] T. Bollerslev, R. Y. Chou, and K. F. Kroner. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1-2):5–59, 1992.

[BCST12] M. J. Brennan, T. Chordia, A. Subrahmanyam, and Q. Tong. Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics*, 105(3):523–541, 2012.

[BDB17] P. Blanc, J. Donier, and J.-P. Bouchaud. Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17(2):171–188, 2017.

[BDHM13a] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[BDHM13b] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

[BDM12] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.

[Bel17] M. Bellia. High-frequency market making: Liquidity provision, adverse selection, and competition. *Working paper*, 2017.

[Ber05] L. Bergomi. Smile dynamics II. *Risk*, 18:67–73, 2005.

[BFG16] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[BFG+17] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper. Short-time near-the-money skew in rough fractional volatility models. *arXiv preprint arXiv:1703.05132*, 2017.

[BH04] L. Bauwens and N. Hautsch. Dynamic latent factor models for intensity processes. *CORE Discussion Paper*, 2004.

[BJM16] E. Bacry, T. Jaisson, and J.-F. Muzy. Estimation of slowly decreasing Hawkes kernels: Application to high frequency order book modelling. *Quantitative Finance*, 16(8):1179–1201, 2016.

[Bla76] F. Black. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, pages 171–181, 1976.

[BLP16] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Decoupling the short-and long-term behavior of stochastic volatility. *Available at SSRN 2846756*, 2016.

[BLP17] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

[BLT06] T. Bollerslev, J. Litvinova, and G. Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4(3):353–384, 2006.

[BM13] M. Brolley and K. Malinova. Informed trading and maker-taker fees in a low-latency limit order market. *Working paper*, 2013.

[BMM15] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[Bow07] C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[BP03] J.-P. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: from statistical physics to risk management.* Cambridge university press, 2003.

[BP09] M. K. Brunnermeier and L. H. Pedersen. Market liquidity and funding liquidity. *Review of Financial studies*, 22(6):2201–2238, 2009.

[BW00] G. Bekaert and G. Wu. Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13(1):1–42, 2000.

[CB14] R. Chicheportiche and J.-P. Bouchaud. The fine-structure of volatility feedback i: Multi-scale self-reflexivity. *Physica A: Statistical Mechanics and its Applications*, 410:174–195, 2014.

[CC98] P. Carmona and L. Coutin. Fractional Brownian motion and the Markov property. *ECP*, 3:95–107, 1998.

[CCM00] P. Carmona, L. Coutin, and G. Montseny. Approximation of some Gaussian processes. *Stat. Inference Stoch. Process.*, 3(1-2):161–171, 2000. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).

[CDDM05] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. A point process approach to value-at-risk estimation. *Quantitative Finance*, 5(2):227–234, 2005.

[CF12] J.-E. Colliard and T. Foucault. Trading fees and efficiency in limit order markets. *The Review of Financial Studies*, 25(11):3389–3421, 2012.

[CGP18] G. Callegaro, M. Grasselli, and G. Pagès. Rough volatility is not so tough: fast hybrid schemes for fractional Riccati equations. *Working paper*, 2018.

[CH92] J. Y. Campbell and L. Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318, 1992.

[Chr82] A. A. Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432, 1982.

[CK93] J. Cvitanić and I. Karatzas. Hedging contingent claims with constrained portfolios. *The Annals of Applied Probability*, pages 652–681, 1993.

[CM99] P. Carr and D. Madan. Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

[Cor00] V. Corradi. Reconsidering the continuous time limit of the GARCH (1, 1) process. *Journal of Econometrics*, 96(1):145–153, 2000.

[CPT15] J. Cvitanić, D. Possamaï, and N. Touzi. Dynamic programming approach to principal-agent problems. *arXiv preprint arXiv:1510.07111*, 2015.

[CT18] C. Cuchiero and J. Teichmann. Generalized feller processes and markovian lifts of stochastic volterra processes: the affine case. *arXiv preprint arXiv:1804.10450*, 2018.

[CZ13] J. Cvitanic and J. Zhang. Contract theory in continuous-time methods, 2013.

[DF99] K. Diethelm and A. D. Freed. The fracpece subroutine for the numerical solution of differential equations of fractional order. In *Forschung und Wissenschaftliches Rechnen 1998*, pages 57–71. Gesellschaft für Wisseschaftliche Datenverarbeitung Gottingen, Germany, 1999.

[DFF02] K. Diethelm, N. J. Ford, and A. D. Freed. A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dynamics*, 29(1-4):3–22, 2002.

[DFF04] K. Diethelm, N. J. Ford, and A. D. Freed. Detailed error analysis for a fractional Adams method. *Numerical algorithms*, 36(1):31–52, 2004.

[DFS03] D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *Ann. Appl. Probab.*, 13(3):984–1053, 2003.

[DGE93] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

[DR15] K. Dayri and M. Rosenbaum. Large tick assets: implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 1(01):1550003, 2015.

[Dua97] J.-C. Duan. Augmented GARCH(p, q) process and its diffusion limit. *Journal of Econometrics*, 79(1):97–127, 1997.

[Dwa69] M. Dwass. The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6(3):682–686, 1969.

[DY02] A. A. Dragulescu and V. M. Yakovenko. Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative finance*, 2(6):443–453, 2002.

[EGG10] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and convergence*. 1986.

[ELL11] P. Embrechts, T. Liniger, and L. Lin. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[EN93] R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1749–1778, 1993.

[FGP17] P. Friz, S. Gerhold, and A. Pinter. Option pricing in the moderate deviations regime. *Mathematical Finance*, 2017.

[FJL12] M. Forde, A. Jacquier, and R. Lee. The small-time smile and term structure of implied volatility under the Heston model. *SIAM Journal on Financial Mathematics*, 3(1):690–708, 2012.

[FKK13] T. Foucault, O. Kadan, and E. Kandel. Liquidity cycles and make/take fees in electronic markets. *The Journal of Finance*, 68(1):299–341, 2013.

[FS15] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, (ahead-of-print):1–22, 2015.

[FSS87] K. R. French, G. W. Schwert, and R. F. Stambaugh. Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29, 1987.

[Fuk11] M. Fukasawa. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[Fuk12] M. Fukasawa. The normalizing transformation of the implied volatility smile. *Mathematical Finance*, 22(4):753–762, 2012.

[Fuk17] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

[FW00] S. Figlewski and X. Wang. Is the leverage effect a leverage effect? *Available at SSRN 256109*, 2000.

[FZ17] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1):114–145, 2017.

[Gat11] J. Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[GJR14] H. Guennoun, A. Jacquier, and P. Roome. Asymptotic behaviour of the fractional Heston model. *Available at SSRN 2531468*, 2014.

[GJR18] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[GKR18] J. Gatheral and M. Keller-Ressel. Affine forward variance models. *arXiv preprint arXiv:1801.06416*, 2018.

[GLFT13] O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, 7(4):1–31, 2013.

[GLS90] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra integral and functional equations*, volume 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[GO10] J. Gatheral and R. C. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2):249–283, 2010.

[GR18] J. Gatheral and R. Radoičić. Rational approximation of the rough Heston solution. *SSRN*, 2018.

[GS17] J. Garnier and K. Sølna. Correction to black–scholes formula due to fractional stochastic volatility. *SIAM Journal on Financial Mathematics*, 8(1):560–588, 2017.

[Har13] L. Harris. Maker-taker pricing effects on market quotations. *USC Marshall School of Business*, 2013.

[Haw71] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.

[HBB13] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.

[Hes93] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

[HKLW02] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *The Best of Wilmott*, page 249, 2002.

[HM87] B. Hölmstrom and P. Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, 55:303–328, 1987.

[HMS11] H. J. Haubold, A. M. Mathai, and R. K. Saxena. Mittag-leffler functions and their applications. *Journal of Applied Mathematics*, 2011.

[HO74] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[Höl79] B. Hölmstrom. Moral hazard and observability. *The Bell Journal of Economics*, 10:74–91, 1979.

[HS81] T. Ho and H. R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, 1981.

[HS06] T. Hendershott and M. S. Seasholes. Market maker inventories and stock prices. *Available at SSRN 890860*, 2006.

[HS15] P. Harms and D. Stefanovits. Affine representations of fractional processes with applications in mathematical finance. *arXiv preprint arXiv:1510.04061*, 2015.

[HW87] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300, 1987.

[Itk05] A. Itkin. Pricing options with VG model using FFT. *arXiv preprint physics/0503137*, 2005.

[JKWW11] A. Janek, T. Kluge, R. Weron, and U. Wystup. FX smile in the Heston model. In *Statistical Tools for Finance and Insurance*, pages 133–162. Springer, 2011.

[JMP89] A. Jakubowski, J. Mémin, and G. Pages. Convergence en loi des suites d'intégrales stochastiques sur l'espace 1 de skorokhod. *Probability Theory and Related Fields*, 81(1):111–137, 1989.

[JR13] A. Jacquier and P. Roome. The small-maturity Heston forward smile. *SIAM Journal on Financial Mathematics*, 4(1):831–856, 2013.

[JR15] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

[JR16a] A. Jacquier and P. Roome. Large-maturity regimes of the Heston forward smile. *Stochastic Processes and their Applications*, 126(4):1087–1123, 2016.

[JR16b] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed hawkes processes. *The Annals of Applied Probability*, (To appear), 2016.

[JR18] P. Jusselin and M. Rosenbaum. No-arbitrage implies power-law market impact and rough volatility. *arXiv preprint arXiv:1805.07134*, 2018.

[JS13] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2013.

[KJ05] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott magazine*, pages 94–103, September 2005.

[KP91] T. G. Kurtz and P. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, pages 1035–1070, 1991.

[KT+03] N. Kunitomo, A. Takahashi, et al. On validity of the asymptotic expansion approach in contingent claim analysis. *The Annals of Applied Probability*, 13(3):914–952, 2003.

[Lew01] A. L. Lewis. A simple option formula for general jump-diffusion and other exponential lévy processes. *Available at SSRN 282110*, 2001.

[Lin09] A. M. Lindner. Continuous time approximations to garch and stochastic volatility models. In *Handbook of Financial Time Series*, pages 481–496. Springer, 2009.

[LL13] C.-A. Lehalle and S. Laruelle. *Market microstructure in practice.* World Scientific, 2013.

[LMPR18] G. Livieri, S. Mouti, A. Pallavicini, and M. Rosenbaum. Rough volatility: evidence from option prices. *IISE Transactions*, (just-accepted):1–21, 2018.

[LT09] C. Li and C. Tao. On the fractional Adams method. *Computers & Mathematics with Applications*, 58(8):1573–1588, 2009.

[Lut10] M. Lutat. The effect of maker-taker pricing on market liquidity in electronic trading systems–empirical evidence from european equity trading. *Working paper*, 2010.

[LZ03] P. E. Lynch and G. O. Zumbach. Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, 3(4):320–331, 2003.

[Mai] F. Mainardi. On some properties of the Mittag-Leffler function. *arXiv preprint arXiv:1305.0161.*

[Man97] B. B. Mandelbrot. The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pages 371–418. Springer, 1997.

[Men13] A. J. Menkveld. High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740, 2013.

[MH08] A. M. Mathai and H. J. Haubold. *Special functions for applied scientists.* Springer, 2008.

[Mir74] J. Mirrlees. Notes on welfare economics, information and uncertainty. *Essays on economic behavior under uncertainty*, pages 243–261, 1974.

[MP15a] K. Malinova and A. Park. Subsidizing liquidity: The impact of make/take fees on market quality. *The Journal of Finance*, 70(2):509–536, 2015.

[MP15b] A. Mazzon and A. Pascucci. The forward smile in local-stochastic volatility models. *Available at SSRN 2560300*, 2015.

[MRR97] A. Madhavan, M. Richardson, and M. Roomans. Why do security prices change? a transaction-level analysis of nyse stocks. *The Review of Financial Studies*, 10(4):1035–1064, 1997.

[MS07] A. Medvedev and O. Scaillet. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies*, 20(2):427–459, 2007.

[MS15] L. Mytnik and T. S. Salisbury. Uniqueness for volterra-type stochastic integral equations. *arXiv preprint arXiv:1502.05513*, 2015.

[MSLR17] N. Megarbane, P. Saliba, C.-A. Lehalle, and M. Rosenbaum. The behaviour of high-frequency traders under different market stress scenarios. *To appear in Market Microstructure and Liquidity*, 2017.

[Mur11] A. A. Muravlëv. Representation of fractal Brownian motion in terms of an infinite-dimensional Ornstein-Uhlenbeck process. *Uspekhi Mat. Nauk*, 66(2(398)):235–236, 2011.

[Nel90] D. B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1):7–38, 1990.

[Nel91] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

[Nev72] J. Neveu. Martingales à temps discret. 1972.

[NS16] A. Neuenkirch and T. Shalaiko. The order barrier for strong approximation of rough volatility models. *arXiv preprint arXiv:1606.03854*, 2016.

[NUZ88] D. Nualart, A. Ustunel, and M. Zakai*. On the moments of a multiple wiener-itô integral and the space induced by the polynomials of the integral. *Stochastics*, 25(4):233–340, 1988.

[Osa07] Y. Osajima. The asymptotic expansion formula of implied volatility for dynamic sabr model and fx hybrid model. 2007.

[Poo09] S.-H. Poon. The Heston option pricing model. *Unpublished Draft*, 2009.

[RP12] R. Riordan and A. Park. Maker/taker pricing and high frequency trading. *Foresight report, UK Government Office for Science*, 2012.

[RR12] M. J. Rodríguez and E. Ruiz. Revisiting several popular garch models with leverage effect: Differences and similarities. *Journal of Financial Econometrics*, 10(4):637–668, 2012.

[RY13] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[San08] Y. Sannikov. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984, 2008.

[Sch10] M. Schmelzle. Option pricing formulae using Fourier transform: Theory and application. *Preprint, http://pfadintegral. com*, 2010.

[S.E10] S.E.C. Concept release on equity market structure. *Release No. 34-61358; File No. S7-02-10.*, 2010.

[SKM93] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional integrals and derivatives*, volume 1993. Theory and Applications, Gordon and Breach, Yverdon, 1993.

[Sok13] A. Sokol. Optimal Novikov-type criteria for local martingales with jumps. *Electronic Communications in Probability*, 18:1–8, 2013.

[SS91] E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4(4):727–752, 1991.

[TT12] R. Tayal and S. Thomas. Measuring and explaining the asymmetry of liquidity. *Available at SSRN 2239492*, 2012.

[Ver12] M. Veraar. The stochastic fubini theorem revisited. *Stochastics An International Journal of Probability and Stochastic Processes*, 84(4):543–551, 2012.

[WBK+08] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[WM14] C. D. Wang and P. A. Mykland. The estimation of leverage effect with high-frequency data. *Journal of the American Statistical Association*, 109(505):197–215, 2014.

[WX02] G. Wu and Z. Xiao. A generalized partially linear model of asymmetric volatility. *Journal of Empirical Finance*, 9(3):287–319, 2002.

[Yos92] N. Yoshida. Asymptotic expansion for statistics related to small diffusions. *Journal of the Japan Statistical Society, Japanese Issue*, 22(2):139–159, 1992.

[YW71] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.

[Zak94] J.-M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994.

[ZL01] G. Zumbach and P. Lynch. Heterogeneous volatility cascade in financial markets. *Physica A: Statistical Mechanics and its Applications*, 298(3-4):521–529, 2001.

[Zum04] G. Zumbach. Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4(1):70–86, 2004.

[Zum09] G. Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.