



Catholic University of Louvain
Institute of Statistics, Biostatistics and Actuarial Sciences

A Random Walk in Illiquidity Modeling

Rough volatility, subdiffusions, fractional Hawkes, optimal
liquidation and deep learning

Jean-Loup DUPRET

A thesis submitted to the
Catholic University of Louvain
in partial fulfillment of the
requirements for the degree of
DOCTOR OF SCIENCES

Thesis Committee:

Prof. **Patrick CHERIDITO**
Prof. **Donatien HAINAUT**
Prof. **Christian HAFNER**
Prof. **Michèle VANMAELE**
Prof. **Frédéric VRINS**

ETH Zurich
Catholic University of Louvain
Catholic University of Louvain
Ghent University
Catholic University of Louvain

Supervisor
President

Secretary

June 2024

*Begin at the beginning, and go on till you come to the end.
Then, stop.*

Lewis CARROLL, *Alice's Adventures in Wonderland*.

Acknowledgements

Cette thèse n'aurait pu voir le jour sans l'apport précieux et la collaboration étroite de nombreuses personnes. Chaque étape de cette aventure a été marquée par les contributions variées et essentielles de mes proches et collègues sans lesquels cette thèse n'aurait pas pu se concrétiser.

Je souhaite en premier lieu exprimer ma gratitude à mon promoteur, Prof. Donatien Hainaut, pour ces quatre années de travail passionnant. Je lui suis infiniment reconnaissant, non seulement pour son encadrement scientifique, mais aussi pour sa confiance, son soutien et la liberté accordée en termes de sujets de recherche.

Je tiens également à exprimer ma gratitude à tous les membres de mon jury de thèse. Tout d'abord, Prof. Michèle Vanmaele et Prof. Frédéric Vrins pour avoir fait partie de mon comité d'encadrement et pour les nombreuses remarques, notamment lors de l'épreuve de confirmation et la défense privée, qui ont contribué significativement à l'amélioration de ma recherche. Je remercie également le Président du Jury, Prof. Christian Hafner, pour les nombreux échanges, sa bienveillance et sa disponibilité. Finalement, je tiens particulièrement à remercier Prof. Patrick Cheridito pour m'avoir offert l'opportunité de faire ce séjour de recherche à l'ETH Zurich et de m'accorder sa confiance pour la suite de mon parcours académique.

Je tiens tout spécialement à remercier mon promoteur de mémoire Dr. Jérôme Barbarin pour avoir éveillé en moi la passion de la recherche et des mathématiques financières.

Cette thèse n'aurait pas été possible sans le soutien administratif, logistique et calorifique (au vu des nombreuses sucreries disponibles en permanence à la cafétéria) composés de Maguy, Nadja, Nancy, Sophie et Tatiana. Merci pour votre organisation, votre disponibilité et votre bonne humeur. Je souhaiterais aussi remercier les différents professeurs de l'ISBA ainsi que les membres du SMCS avec lesquels j'ai pu collaborer d'une façon ou d'une autre.

Je souhaiterais ensuite remercier tous les collègues doctorants et post-doctorants avec qui j'ai pu échanger au quotidien durant ces quatre années et qui ont façonné à leur manière ma thèse de doctorat. Je tiens en premier lieu à remercier mes collègues de bureaux pour avoir rendu l'atmosphère de travail aussi chaleureuse : Charlotte, Chikéola, Keivan, Patricia et Vanessa. De plus, j'espère n'oublier personne, parmi les nouveaux comme les anciens, en remerciant : Aigerim, Alexandre, Anas, Antoine, Aurèle, Benjamin, Charles, Chikéola, Ensyieh, Fanny, Gabriel, Gilles, Hortense, Hugo, John-John, Keivan, Lara, Lise, Madeline, Mathilde, Mengxue, Morine, Nathalie, Oussama, Quentin, Stefka, Stéphane. Merci d'avoir contribué à rendre cette thèse moins solitaire, beaucoup plus conviviale et d'avoir significativement amélioré mes talents au kicker ($p\text{-val} < 0.001$). Je remercie en particulier Édouard pour l'aide précieuse apportée en fin de thèse.

Je n'aurais probablement jamais pu commencer cette thèse sans le soutien de ma famille, et en particulier de mes parents. Avec le recul, je me rends compte de l'importance de tous les sacrifices que vous avez faits et du temps que vous avez pris pour me soutenir et me permettre de m'épanouir. Je tiens également à remercier tout particulièrement mes proches et amis. Merci pour votre soutien, votre présence et vos encouragements.

Merci également à ceux qui ne m'en voudront pas de ne pas les avoir cités.

Finalement, je remercie la personne qui m'a été le plus importante ces quatre dernières années. Julie, je te remercie énormément pour tout ce que tu m'apportes au quotidien. Même si je doute que tu la lises un jour entièrement, c'est à toi que je dédie cette thèse.

Contents

Introduction	1
I Rough volatility modeling	9
1 Impact of rough stochastic volatility models on long-term life insurance pricing	11
1.1 Introduction	12
1.2 Life insurance contract	13
1.3 Key features of the RFSV model	14
1.3.1 rBergomi model	16
1.3.2 Rough Heston	17
1.3.3 Long-term behavior of rBergomi and rough Heston models	19
1.4 Calibration	20
1.4.1 Market data	21
1.4.2 Heston and Bates models	22
1.4.3 Rough Heston calibration	25
1.4.4 rBergomi Calibration	28
1.5 Volatility surface fit	30
1.5.1 Comparison of the fit	32
1.6 Simulation and discretization methodology	33
1.7 Conclusion of the fit	37
1.8 Life insurance contract valuation	38
1.8.1 Valuation	38
1.9 Conclusion	42
1.A SSVI forward variance curve	44
1.B Calibration and valuation based on other forward variance curves	46
1.B.1 Constant forward variance curve	46
1.B.2 Market forward variance curve	48
1.B.3 Forward Heston model	49

1.C	Figures	51
1.C.1	Risk-free rates	51
1.C.2	Conditional variance process	51
2	Portfolio insurance under rough volatility and Volterra processes	53
2.1	Introduction	54
2.2	Model setup under rough volatility	55
2.3	CPPI in the rough Heston model	58
2.4	Estimation of the rough Heston model	61
2.4.1	Filtering	61
2.4.2	Econometric estimation	63
2.5	Other models	65
2.5.1	Heston model	65
2.5.2	Affine Volterra processes	67
2.5.3	Black and Scholes model	70
2.5.4	(Fractional) Riccati equations	71
2.6	Moments	72
2.7	Risk Measure and probability density	74
2.7.1	Application	75
2.8	Discrete-time CPPI	78
2.9	Conclusion	80
2.A	Fractional Calculus	82
2.B	Proof of Theorem 2.1	82
2.C	Algorithms	84
2.D	Proof of Theorem 2.2	85
2.E	Proof of Corollary 2.3	86
2.F	Fast Fourier Transform	87
II	Illiquidity modeling with fractional processes	89
3	A subdiffusive stochastic volatility jump model	91
3.1	Introduction	91
3.2	A Fokker-Planck equation for jump-diffusion models	95
3.3	Subordinators	97
3.4	A Fokker-Planck equation for the subdiffusive SVJ (SSVJ) model . .	102
3.5	Change of measure	108
3.5.1	Change of measure for the classical SVJ model	109
3.5.2	Change of measure for the SSVJ model	110
3.6	An estimation procedure	112
3.6.1	Filtering	112
3.6.2	Econometric estimation	117
3.7	Risk management and option pricing	119

3.7.1	Risk management	124
3.7.2	Pricing	125
3.8	Conclusion	129
3.A	Proof of Proposition 3.4	131
3.B	Second simulated example	132
3.C	Existence of SSVJ moments	132
3.D	Proof of Proposition 3.1 (iii).	133
4	A fractional Hawkes process for illiquidity modeling	135
4.1	Introduction	136
4.2	Hawkes processes	138
4.2.1	The modified Mittag-Leffler fractional Hawkes process	140
4.2.2	Standard Hawkes processes	141
4.2.3	Fractional Hawkes processes	144
4.2.4	Comparison	145
4.3	The modified ML fractional Hawkes process	146
4.3.1	Properties	146
4.3.2	Spectral representation	147
4.3.3	Discretization and conditional transform	149
4.3.4	A multivariate extension	152
4.4	Illiquidity modeling	154
4.4.1	Amihud measure	154
4.4.2	Estimation procedure	156
4.4.3	Characteristic function of the log-Amihud measure	162
4.5	Risk management on illiquidity	163
4.6	Conclusion	165
4.A	Data section	166
4.B	Proofs	168
4.B.1	Proof of Proposition 4.1.	168
4.B.2	Proof of Proposition 4.2.	169
4.B.3	Proof of Proposition 4.4	170
4.B.4	Proof of Proposition 4.5.	170
4.B.5	Proof of Corollary 4.1.	172
4.B.6	Proof of Corollary 4.2.	173
4.B.7	Poof of Proposition 4.6	174
III	Optimal control in finance	177
5	Optimal liquidation under indirect price impact with propagator	179
5.1	Introduction	180
5.2	The indirect price impact model with propagator	185
5.2.1	Temporary price impact kernel	188
5.2.2	Permanent price impact kernel	198

5.2.3	Transient impact function	204
5.3	Indirect optimal liquidation with limit and market orders	209
5.4	Conclusion	214
5.A	Dirac's delta propagator	216
5.B	Alternative risk-aversion penalty	217
6	Deep learning for high-dimensional stochastic optimal control without explicit solution	219
6.1	Introduction	220
6.2	Stochastic optimal control	222
6.2.1	Explicit optimal control	225
6.2.2	Implicit optimal control	227
6.3	Deep learning for implicit CT stochastic optimal control	231
6.3.1	Standard PINN	231
6.3.2	Sequential PINN	232
6.4	Neural network implementation	237
6.5	A comparative example	239
6.6	A practical example in mathematical finance	244
6.6.1	Multi-asset optimal liquidation	248
6.7	Conclusion	251
6.A	Proof of Theorem 6.3	252
	Conclusion and extensions	257
	References	261

Introduction

Liquidity is a fundamental property of a well-functioning market and lack of liquidity is generally at the heart of many financial crises and disasters, such as the financial crisis of 2007–2008 and the LTCM crisis of 1998. Simply stated, liquidity refers to the efficiency or ease with which an asset or security can be sold in financial markets without affecting its market price. More specifically, liquidity can be thought of as an exogenous measure of the added costs per transaction associated with asset trading, either from transaction costs or market impact. The primary source of illiquidity therefore comes from exogenous transaction costs such as brokerage fees, order-processing costs, or transaction taxes. Every time a security is traded, the buyer and/or seller indeed incurs a transaction cost. The second main source of illiquidity then comes from market impact, which reflects the impact of order flow on price – the discount that a seller concedes or the premium that a buyer pays when executing a market order – and which results from demand pressure and inventory costs. Demand pressure arises because not all agents are present in the market at all times, which means that if an agent needs to sell (or buy) a security quickly, then the natural buyers (or sellers) may not be immediately available. As a result, the seller may sell to a market maker who buys in anticipation of being able to later lay off the position. The market maker, being exposed to the risk of price changes while he holds the asset in inventory, must be compensated for this risk – a compensation that imposes an (inventory) cost on the seller. For standard-size transactions, the price impact is the bid–ask spread, whereas larger excess demand induces a greater impact on prices, reflecting a likely action of informed traders as explained in Easley and O’hara (1987). Because market makers cannot distinguish between order flow that is generated by informed traders and that generated by liquidity (noise) traders, they set prices that are an increasing function of the imbalance in the order flow which may indicate informed trading, see Kyle (1985). This creates a positive relationship between the order flow (or transaction volume) and price change, precisely referred in the literature to as price/market impact.

Liquidity is always tied to the temporal granularity at which it is analyzed. In

mathematical finance, when modeling the price of an asset, we indeed typically distinguish between at least three different time scales corresponding to the level of granularity at which the trades and their liquidity are studied in the market: high-frequency (fraction of seconds), intermediate or intraday time scale (between seconds and hours) and low-frequency (from days to years for very large time scales). These three temporalities and their connections are the subject of the three parts constituting this thesis.

At the high-frequency level, high-frequency price models aim at reproducing some statistical observations made on market data such as the autocorrelation in the signs of trades, the market liquidity, the volatility clustering effect, the high-frequency resilience of the price, as well as the discrete nature of time trade arrivals and of price variation. At very high frequencies, one then has to describe the limit order book (LOB) dynamics, or a part of it. Such models have been proposed by Abergel and Jedidi (2013), Cont and De Larrard (2013), Gareche et al. (2013), Huang et al. (2015), Da Fonseca and Zaatour (2014) and especially in Bacry and Muzy (2014) with a tick-by-tick price model ruled by Hawkes processes, which describes jointly the order flow and the price dynamics and which reproduces numerous empirical properties of high-frequency markets. Moreover, these high-frequency price models also aim at obtaining low-frequency asymptotics that are consistent with continuous diffusions. Asymptotic and low-frequency behaviour of such models have indeed been recently investigated by Bacry and Muzy (2014) and Jaisson and Rosenbaum (2016). In particular, we are especially interested in the rough behavior of volatility as it is shown in El Euch et al. (2018) that this rough volatility can be explained from the microscopic high-frequency interactions between agents. El Euch et al. (2018) indeed build a tick-by-tick price model encoding the microscopic stylized facts of market liquidity in the context of high frequency trading, using unstable Hawkes processes. By studying the asymptotic behavior of this microscopic price dynamic in the long run, they obtain in the low-frequency limit a rough version of the Heston model. The first part of the thesis is more empirical and therefore aims at studying rough volatility (and in particular the rough Heston model) as it arises from the modeling of market liquidity at high-frequency. We hence first expand in Chapter 1 these rough volatility models at larger time scales by pricing long-term life insurance contracts, thereby contributing to bridging the gap between high-frequency liquidity modeling and asset pricing at very large time scales. We then consider in Chapter 2 the rough Heston model (and more generally affine Volterra processes) in the context of portfolio insurance strategies.

On the other end, when dealing with low frequencies, the price can often be well approximated by a diffusive process. Ever since the advent of the Black and Scholes (1973) option pricing formula, the study of asset modeling based on diffusive processes has become a central preoccupation for both academics and practitioners, see the review of Musiela and Rutkowski (2006) or Bergomi (2015). Additionally,

various econometric models in discrete time have also been explored in the literature to directly model the illiquidity process of these assets at such lower time scales, see Gökay et al. (2011), Brennan et al. (2013) or Amihud (2015). Inspired by both these diffusive and econometric models, the second part of this thesis demonstrates how the illiquidity process associated with such low-frequency observations of market prices can be modeled directly using diffusions and fractional processes. In the third chapter, we indeed introduce a new approach for modeling the illiquidity process of thinly-traded assets using an inverse subordinator. This implies focusing on subdiffusions rather than diffusive processes to represent the price of illiquid assets in financial markets. Our focus then shifts in the fourth chapter to the Amihud illiquidity measure, which is widely used for measuring the illiquidity process of stocks and indices on a daily time scale. Again, we directly model this Amihud measure of illiquidity using a mean-reverting jump diffusion, with the jumps following a new fractional Hawkes process.

In between the high-frequency and low-frequency models, the third part of the thesis investigates price impact models that consider an intraday intermediate (or mesoscopic) time scale, somewhere between seconds and hours. They focus on describing the price/market impact of transactions, one of the primary sources of market illiquidity arising from the net order flow imbalance, as discussed above. Their goal is to be more tractable than high-frequency models and to bring quantitative results on practical issues such as optimal liquidation strategies. Optimal liquidation indeed refers to the situation in which a trader plans to sell (or buy) a large number of units of a risky asset before a predetermined time horizon. Since the size of the order is large, this trader may find it more optimal to work the order in several smaller slices to minimize her impact on prices by trading during times of higher liquidity and taking advantage of the resilience of the supply and demand of the asset. On the other hand, delaying the orders for too long increases the exposure to other market risks. The goal is hence to find the right balance between liquidity risk (*i.e.* the price/market impact) from trading quicker and the other market risks from taking longer to complete all trades. We refer to Cartea et al. (2015) and Gökay et al. (2011) for a more thorough introduction on the subject. Chapter 5 hence proposes a new jump-dependent price impact model with a stochastic liquidity factor, aiming at best reproducing the empirical effects of market orders on liquidity and transaction prices. We then address the associated optimal execution problem for the liquidating agent based on this novel price impact model. Finally, we develop in Chapter 6 a new deep-learning algorithm for solving complex high-dimensional stochastic optimal control problems in continuous-time. Specifically, we focus in this chapter on optimal liquidation problems without explicit solution involving a large number of co-integrated assets.

Outline of the thesis

The three parts of the thesis contain six chapters which are self-contained research articles that can be read independently from each other. The outline of the dissertation is as follows.

I. Rough volatility modeling: Chapters 1 and 2

Rough volatility has been first introduced in Gatheral et al. (2018) with the Rough Fractional Stochastic Volatility (RFSV) model, which is based on a fractional Ornstein-Uhlenbeck process with a Hurst index below $1/2$, in contrast to the long-memory fractional Ornstein-Uhlenbeck model of Comte and Renault (1998) with Hurst index above $1/2$. The RFSV model has been shown to be remarkably consistent with financial time series of past volatility data as well as with the observed implied volatility surface. However, due to the need of Monte-Carlo simulations, this RFSV does not provide rapidly option prices, nor an effective calibration procedure. Two tractable implementations were then derived from the RFSV with the rBergomi model of Bayer et al. (2016) and the rough Heston model of El Euch et al. (2019). As explained in the general introduction, El Euch et al. (2018) demonstrate how this rough Heston model arises as the low-frequency limit from the modeling of market liquidity at high-frequency. We hence contribute with Chapter 1 to bridging this gap between high-frequency market liquidity and long-term asset/insurance pricing. This first empirical chapter indeed shows how to expand these two rough volatility models (and more generally any forward-variance models) at larger time scales, analyzes their implications for the pricing of long-term life insurance contracts and explains why they provide a more accurate fair value of such long-term contracts. In particular, we highlight and study the long-term properties of these two rough volatility models and compare them with standard stochastic volatility models, such as the Heston and Bates models. For the rough Heston model, we manage to build a highly consistent calibration and pricing methodology based on a stable and arbitrage-free regime for the volatility at large maturity. This ensures a reasonable behavior of the model in the long run. Regarding the rBergomi model, we show that it does not exhibit a realistic long-term volatility with extremely large swings at large time scales and that it is not fast enough for calibration purposes, in contrast to the rough Heston model which is highly tractable. Therefore, compared to standard stochastic volatility models, the rough Heston model efficiently provides a more accurate fair value of long-term life insurance contracts embedding path-dependent options while being highly consistent with historical and risk-neutral data.

Chapter 2 then studies in the context of portfolio insurance the class of affine Volterra processes which encompasses this rough Heston model (as well as the classical Heston model). The aim of this chapter is hence to revisit and generalize

the constant proportion portfolio insurance (CPPI) under affine Volterra processes. Indeed, existing simulation-based methods for CPPI do not apply easily to this class of processes. We instead propose an approach based on the characteristic function of the log-cushion which appears to be more consistent, stable and particularly efficient in the case of affine Volterra processes, compared with existing simulation techniques. Using such approach, we describe in this chapter several properties of CPPI which naturally result from the form of the log-cushion's characteristic function under affine Volterra processes. This allows to consider more realistic dynamics for the underlying risky asset in the context of CPPI and hence build portfolio strategies that are more consistent with financial data. In particular, we are especially interested in the rough Heston model as it arises from the high-frequency modeling of market liquidity, and we provide a new estimation procedure for its parameters based on the Particle Monte-Carlo Markov Chain (PMCMC) algorithm. Finally, the last part of the chapter investigates gap risk in discrete-time CPPI, which arises from the liquidity constraints and the transaction costs discussed in the introduction. Therefore, in Chapter 2, we aim to study more accurately the true properties of the CPPI strategy and to better manage the (liquidity) risk associated with it.

II. Illiquidity modeling with fractional processes: Chapters 3 and 4

Our focus then shifts in the second part of the thesis to directly modeling the illiquidity process associated with low-frequency observations of asset prices using fractional processes. Chapter 3 begins by investigating small capitalization stocks, which tend to exhibit motionless periods in their share price corresponding to a lack of liquidity associated with these illiquid assets. This occurs when a stock stops being exchanged by the investors during several days, weeks or months. Subdiffusions (or *fractional diffusions*) appear as good candidates for modeling the price of such illiquid financial assets since they are obtained from subordination, where a standard diffusion is time-changed by an inverse subordinator representing the illiquidity process of the associated stock price. Existing subdiffusive models of asset prices are indeed able to capture the motionless periods in the quotes of thinly-traded assets. However, they fail at reproducing simultaneously the jumps and the time-varying random volatility observed in their sample paths. The aim of this chapter is hence to propose a new model of subdiffusive asset prices reproducing the main characteristics exhibited in illiquid markets. This is done by considering a stochastic volatility jump model, time-changed by an inverse subordinator driving the (il)liquidity process of the related asset. We then derive the forward fractional partial differential equations (PDE) governing the probability density function of the introduced model and we prove that it leads to an arbitrage-free and incomplete market under a suitable change of measure. By proposing a new procedure for estimating the model parameters and using a series expansion for solving numerically the obtained fractional PDE, we are able to price various European-type

derivatives on illiquid assets and to depart from the common Markovian valuation setup. This way, we show that the introduced subdiffusive stochastic volatility jump model yields consistent and reliable results in illiquid markets for thinly-traded assets.

We again consider in Chapter 4 a fractional process for directly modeling the illiquidity process associated with asset prices at low-frequency but focus instead on the Amihud illiquidity measure. Indeed, the Amihud illiquidity measure has proved to be very popular in the economic and financial literature for measuring the illiquidity process of stocks and indices at the daily time scale. None of the existing discrete-time illiquidity models are however adapted for reproducing peaks of illiquidity with long-memory, for risk management and for pricing liquidity-related derivatives. This Chapter 4 therefore proposes a new continuous-time paradigm for modeling illiquidity via a novel fractional Hawkes process in which the intensity process is ruled by a modified Mittag-Leffler excitation function. By considering a mean-reverting jump diffusion for the (log-)Amihud measure with jumps following this modified fractional Hawkes process, we then manage to effectively reproduce the observed peaks of illiquidity in financial markets while introducing long-memory and tractability in the model. This model can then be extended in a multivariate setting so as to study the contagion of up-and-down illiquidity shocks between multiple financial assets. Finally, by deriving the characteristic function satisfied by the Amihud measure in this model, we provide in the last part of the chapter an efficient way to perform risk management on illiquidity with the distribution, moments and risk measures of this Amihud measure. These two Chapters 3 and 4 hence introduce new tools based on fractional processes for a better management of the illiquidity risk in financial markets at the low-frequency level.

III. Optimal control in finance: Chapters 5 and 6

At last, the third part of the thesis delves into stochastic optimal control in finance. We specifically focus on optimal liquidation problems as they are built upon price impact models and thus consider liquidity modeling from an intermediate (intra-day) time scale. Chapter 5 then proposes a new framework of optimal liquidation strategies for a trader seeking to liquidate his large inventory based on a jump-dependent price impact model with propagator. This new jump-dependent price impact model best reproduces the empirical direct and indirect effects of market orders on the transaction price by introducing a stochastic liquidity factor in the form of a trade-driven intensity process. Moreover, different choices of propagators are proposed and their implications in terms of temporary, permanent and transient impacts on the transaction price are discussed. For each choice of such kernels, we formulate the most relevant optimal liquidation problem faced by the trader, derive explicitly the related Hamilton-Jacobi-Bellman equation and solve it numerically. Moreover, we demonstrate how our price impact model can be extended to

incorporate the use of limit orders by the liquidating trader. Therefore, in this Chapter 5, we aim to propose an alternative, realistic and flexible description of the order book's dynamics and its liquidity, thereby contributing to bridging the gap between high-frequency liquidity modeling and optimal execution problems at intraday time scales.

Several of the optimal liquidation strategies discussed in the previous chapter however lead to a high-dimensional optimal control which does not admit an explicit representation. In this case, most of the recent literature on deep learning for solving high-dimensional partial differential equations do not apply, see *e.g.* Han et al. (2018) or Beck et al. (2021). This Chapter 6 therefore introduces the GPI-PINN algorithm, a novel numerical scheme for solving continuous-time stochastic optimal control problems in high dimensions when the optimal control does not admit an explicit solution. Combining Physics-Informed Neural Networks with an Actor-Critic structure based on the Generalized Policy Iteration technique, this successive deep learning algorithm employs two separate neural networks to approximate both the value function and the multidimensional optimal control. This way, the GPI-PINN algorithm manages to achieve a global approximation of the optimal solution across all time and space, which can be evaluated online rapidly. The optimality and convergence of the scheme are demonstrated theoretically and its accuracy and efficacy are shown empirically based on two numerical examples. In particular, we generalize the standard Almgren and Chriss (2001) model arising from optimal liquidation by allowing for a price impact model with fully non-linear temporary and permanent impact functions and by considering a multidimensional setting with numerous co-integrated assets. This indeed leads to a high-dimensional implicit optimal control problem that can be effectively handled by our GPI-PINN algorithm.

Published material

Chapter 1 is based on the paper Dupret et al. (2023), published in the European Actuarial Journal, and extends the results of the Master's thesis Dupret and Barbarin (2020). Chapter 2 is derived from Dupret and Hainaut (2021), published in the International Journal of Theoretical and Applied Finance. Chapter 3 is based on Dupret and Hainaut (2023c), published in Quantitative Finance. Chapters 4 and 5 can be found in Dupret and Hainaut (2023a) and Dupret and Hainaut (2023b) and have been submitted to Mathematics and Financial Economics and to the Journal of Economic Dynamics and Control, respectively. Chapter 6 is still a work in progress.

Part I

Rough volatility modeling

CHAPTER 1

Impact of rough stochastic volatility models on long-term life insurance pricing

The Rough Fractional Stochastic Volatility (RFSV) model of Gatheral et al. (2018) is remarkably consistent with both financial time series of past volatility and with the observed implied volatility surface. Two tractable implementations are derived from the RFSV with the rBergomi model of Bayer et al. (2016) and the rough Heston model of El Euch et al. (2019). We show practically in this chapter how to expand these two rough volatility models (and more generally any forward variance models) at larger time scales, we analyze their implications for the pricing of long-term life insurance contracts and we explain why they provide a more accurate fair value of such long-term contracts. In particular, we highlight and study the long-term properties of these two rough volatility models and compare them with standard stochastic volatility models, such as the Heston and Bates models. For the rough Heston model, we manage to build a highly consistent calibration and pricing methodology based on a stable regime for the volatility at large maturity, which ensures a reasonable behavior of the model in the long run. Regarding the rBergomi model, we show that it does not exhibit a realistic long-term volatility with extremely large swings at large time scales and that it is not fast enough for calibration purposes, unlike the highly tractable rough Heston model. Therefore, compared to standard stochastic volatility models, the rough Heston efficiently provides a more accurate fair value of long-term life insurance contracts embedding path-dependent options while being highly consistent with historical and risk-neutral data. As it is shown in El Euch et al. (2018) that the rough Heston model arises as the low-frequency limit from the modeling of market liquidity at high-frequency, we also contribute with Chapter 1 to bridging the gap between high-frequency liquidity modeling and long-term asset/insurance pricing.

Dupret, J. L., Barbarin, J., & Hainaut, D. (2023). Impact of rough stochastic volatility models on long-term life insurance pricing. *European Actuarial Journal*, 13(1), 235-275.

1.1 Introduction

Standard stochastic volatility (SV) models such as the well-known Heston or Bates models have been introduced long ago to address the shortcomings of the celebrated model of Black & Scholes. However, such models are still limited and cannot reproduce some important empirical facts of the historical volatility and of the observed implied volatility surface.

In their seminal paper, Gatheral et al. (2018) propose a model called "*Rough Fractional Stochastic Volatility*" (RFSV) where the log-volatility process is modeled in terms of a fractional Brownian motion. More precisely, a fractional Ornstein-Uhlenbeck process is used with $H < 1/2$, in contrast with the "*Fractional Stochastic Volatility*" model (FSV) previously introduced by Comte and Renault (1998), where the Hurst index satisfies $H > 1/2$. Gatheral et al. (2018) find a highly consistent model with empirical estimates of the volatility time series. Moreover, the RFSV volatility process is stationary, which ensures reasonableness of the model in the long run. However, due to the need of Monte-Carlo simulations, this RFSV model does not provide rapidly option prices, nor an effective calibration procedure. More tractable pricing models such as the rBergomi of Bayer et al. (2016) or the rough Heston of El Euch et al. (2019) were then derived from the RFSV model to improve the efficiency of model calibration and option pricing. These two models appear to be extremely good at fitting the implied volatility surface while still reproducing the empirical properties of the historical volatility due to their rough sample paths. This first chapter shows practically how to expand these two rough-type models (and more generally any forward variance models) at larger time scales in order to price long-term life insurance contracts and explains why they provide a more accurate value of such contracts.

First, we will show that the stationary property of the volatility in the RFSV model will be lost when considering the rBergomi and rough Heston models. Despite this non-stationary property, we will manage to build for the rough Heston model a stable regime for the volatility in the long-run so as to price consistently long-term life insurance claims. On the contrary, for the rBergomi model, we will show that the generated volatility at large time scales will not be realistic and will tend to exhibit extremely large swings. Moreover, since these long-term properties of the rough Heston and rBergomi models depend heavily on the forward variance curve, we will show how to extrapolate the values of this curve beyond the observed maturities. In particular, we will derive a methodology based on a SSVI parametrization of the volatility surface which allows the forward variance curve to be consistent with the absence of static arbitrage and to be stable for large maturities. We will then verify practically these long-term properties by valuating an equity-linked endowment embedding a cliquet option (using four different models: Heston, Bates, rBergomi and rough Heston). This will allow to highlight the

added-value of using rough-volatility models (and especially the rough Heston) for pricing such long-term life insurance contracts.

The main contribution of this chapter is therefore to propose a calibration and a pricing methodology for long-term life insurance contracts using rough volatility models. We also contribute with this work to highlight and study the divergent long-term properties between rough volatility models and standard SV models. This way, we will be able to explain why these rough volatility models provide a more consistent and accurate fair value of such long-term claims. We will indeed find a significantly different fair value of our equity-linked endowment with models incorporating rough volatility, especially at large time scales with the rough Heston model. In fact, we will see that this rough Heston model exhibits the best long-term properties while being highly tractable. This chapter also aims to compare these models in terms of fit with the observed European option prices and implied volatility surface, which will confirm the high robustness of rough volatility models (and especially of the rough Heston model).

1.2 Life insurance contract

Our aim is thus to price the following equity-linked endowment using rough volatility models and compare them with standard SV models, such as the Heston and Bates models. We thus consider an insurance contract that provides a lump sum payment at maturity T in case of survival, or a payment at the time of death if it occurs before T . We first denote by τ_0 the random residual lifetime of the policyholder at time $t = 0$. We suppose an equity-linked case, where payment amounts depend on the market value of a reference fund F_t and where an amount F_0 is invested in it at $t = 0$. We then introduce a yearly minimal guarantee κ_g for the policyholder and a maximal yearly return κ_m that this policyholder can earn on the fund. The survival benefit is then given by $F_T \mathbb{1}_{\tau_0 \geq T}$ and the death benefit by $F_{\tau_0} \mathbb{1}_{\tau_0 < T}$, where F_t is given by a series of guarantees (cliquet option) for $t \in [0, T]$,

$$F_t = F_0 \prod_{u=1}^{\lfloor t \rfloor} \min \left\{ e^{\kappa_m} ; \max \left\{ 1 + \pi \left(\frac{S_u}{S_{u-1}} - 1 \right) ; e^{\kappa_g} \right\} \right\} \\ \times \min \left\{ e^{\kappa_m(t - \lfloor t \rfloor)} ; \max \left\{ 1 + \pi \left(\frac{S_t}{S_{\lfloor t \rfloor}} - 1 \right) ; e^{\kappa_g(t - \lfloor t \rfloor)} \right\} \right\}. \quad (1.2.1)$$

In the equation above, $S = (S_t)_{t \geq 0}$ is the price process of each fund unit defined on a filtration $(\mathcal{F}_t)_{t \geq 0}$. The cliquet above implicitly assumes yearly resettlements. Moreover, the rate π identifies the portion of yearly return recognized to the policyholder with $\pi \in (0, 1]$.

In this life insurance claim, we will model mortality using the following assumption-

tions. We first introduce the non-homogeneous Poisson process $N = (N_t)_{t \geq 0}$ defined on the filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \geq 0}, \mathbb{P})$ by $N_t := \mathbb{1}_{\tau_0 < t}$ and which is equal to zero as long as the individual is alive and jumps to one at death. We then assume independence between the mortality process N and the price process S . The intensity of N_t is given by the deterministic Makeham force of mortality $\mu_x(t) = c + ab^{x+t}$ for a policyholder aged x at time $t = 0$. Hence, her survival probability at time T , being alive at $t = 0$, is equal to

$${}_T p_x = \mathbb{P}(N_T = 0 | \mathcal{G}_0) = \exp \left(- \int_0^T \mu_x(u) du \right).$$

We now introduce the main characteristics of the RFSV model from Gatheral et al. (2018) and then derive two tractable models from it, the rBergomi and the rough Heston models, that will help us to give a more consistent and accurate market value to this insurance contract.

1.3 Key features of the RFSV model

Gatheral et al. (2018) build the Rough Fractional Stochastic Volatility model (RFSV) with the following SDE under the real-world measure $\mathbb{P}$:

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t, \quad (1.3.1)$$

$$d \log(\sigma_t) = \lambda(\theta - \log(\sigma_t))dt + \nu dB_t^H, \quad (1.3.2)$$

where μ is a drift term, W a standard Brownian motion and B^H is a fractional Brownian motion (fBm) with Hurst exponent H . W and B^H are in general correlated through the constant correlation ρ between W and the Brownian motion driving the Mandelbrot and Van Ness (1968) representation of B^H . Recall that the sample paths of B^H are Hölder-continuous of order β for any $\beta < H$. Similarly, sample paths of B^H are almost surely nowhere Hölder-continuous of order β for any $\beta > H$. Therefore, the larger the Hurst exponent H , the smoother the sample paths and the lower the H , the rougher they are. More precisely, when $H < 1/2$, the sample paths of the fBm are called rough. Since $\log(\sigma_t)$ is driven in equation (1.3.2) by such a fBm, the sample paths of $(\log(\sigma_t))_{t \geq 0}$ are also Hölder-continuous of order β for any $\beta < H$ and exhibit the same regularity behavior as B^H .

Given that equation (1.3.2) models the log-volatility as a fractional Ornstein-Uhlenbeck (fOU) process, we refer to Cheridito et al. (2003) for the derivation of its stationary solution. One important property of the RFSV is that the volatility process $(\sigma_t)_{t \geq 0}$ itself generated by this model is again stationary, which ensures reasonableness of the model at large time scales when pricing long-term claims on the underlying. The cornerstone of Gatheral et al. (2018) is to impose the Hurst exponent H of the fBm B^H to be in $(0, \frac{1}{2})$ and a very long reversion time scale with $\lambda \ll 1/T$. Such conditions aim to generate rough sample paths and to introduce short-range dependence for the volatility process, as explained by Gatheral

et al. (2018). Their findings are extremely consistent at any time scales with empirical statistical properties of the observed volatility time series, as confirmed in Bennedsen et al. (2016). Firstly, Gatheral et al. (2018) show, for $\lambda \ll 1/T$ and $0 < H < 1$, that the log-volatility process $(\log \sigma_t)_{t \geq 0}$ behaves locally as a fBm. Indeed, as $\lambda \rightarrow 0$, they prove that the log-volatility is close to the fBm B^H under the RFSV model (1.3.2). Therefore, this model approximately reproduces the well-known scaling property of the fBm, see Karatzas and Shreve (1998). This scaling property is clearly verified empirically for the discrete historical log-volatility process and allows us to estimate the Hurst index H of the observed volatility process. Many studies (see Bennedsen et al. (2016)) consistently find $H \approx 0.1$, which confirms the rough property of the historical volatility, *i.e.* a Hurst exponent $H < 1/2$. Secondly, neither the autocovariance function $\text{cov}(\sigma_{t+\Delta}, \sigma_t)$ of the volatility process in the RFSV model nor the empirical counterpart of $\text{cov}(\sigma_{t+\Delta}, \sigma_t)$ decay as a power law (cf. *Figure 12* on p. 941 in Gatheral et al. (2018) with $\Delta > 0$). Therefore, these authors claim that historical volatility data are in accordance with the RFSV model, on the contrary to long memory models such as the FSV model of Comte and Renault (1998).

Finally, the RFSV model is not only consistent with empirical statistical properties of the volatility time series but also with the observed implied volatility surface. Indeed, standard stochastic volatility models such as the Hull and White and Heston models do not provide a good fit of the volatility surface (especially for short expirations). Particularly, these models generate a term structure of at-the-money (ATM) volatility skew¹ under a risk-neutral measure $\mathbb{P}^*$ that does not match the observed one for small time to maturity $\tau = T - t$. Instead, empirical studies show that this observed term structure of ATM skew is well approximated by power-law functions of the form $\psi(\tau) \sim \tau^{H-1/2}$ with $0 < H < 1/2$, which can be generated by stochastic volatility models where the log-volatility is driven by fBm with values of the Hurst exponent in $(0, 1/2)$. The explosion of the volatility skew as $\tau \rightarrow 0$ can therefore be modeled using fBm without introducing jumps in our model, as we will confirm in Section 1.5.

Although empirically and theoretically grounded, the RFSV does not allow to rapidly value option prices and therefore to calibrate effectively a pricing model. Indeed, we need Monte-Carlo simulations which are fairly slow, especially for calibration. Therefore, we now show how to adapt the RFSV to obtain more tractable models that can be used to price long-term claims on the underlying. We first analyze in more detail a simple case of the RFSV model when $\lambda = 0$, called the rBergomi model, built upon a forward variance curve. Secondly, we study an extension of the classical Heston model incorporating a rough fractional volatility process. This rough Heston model has the nice property of generating a realistic

¹Recall that the term structure of ATM volatility skew is defined by $\psi(\tau) := \left| \frac{\partial}{\partial k} \hat{\sigma}_t^{imp}(k, \tau) \right|_{k=0}$ where k is the log strike $k := \log S_t/K$ and where $\tau = T - t$.

long-term behavior for the volatility process and of having a characteristic function of the log-stock price in quasi-closed form.

1.3.1 rBergomi model

First, let us denote $V_u = \sigma_u^2$ the instantaneous variance at time u and the forward variance curve (also called variance swap curve) $\xi_t(u) = \mathbb{E}[V_u | \mathcal{F}_t]$, $u \geq t$. We can easily recover V_t from $\xi_t(u)$ as $V_t = \lim_{u \rightarrow t} \xi_t(u)$. Following Bergomi and Guyon (2012), forward variance models are models that can be written as a function of this curve $\xi_t(u)$.

From Bayer et al. (2016), the rBergomi model is obtained from the RFSV model by setting $\lambda = 0$ in the log-volatility dynamic (1.3.2). Under this assumption and from the Mandelbrot & Van Ness representation of the fBm, Bayer et al. (2016) show that the instantaneous variance V_u at time u in the rBergomi model can be rewritten under the physical measure $\mathbb{P}$ as

$$V_u = \mathbb{E}^{\mathbb{P}}[V_u | \mathcal{F}_t] \exp \left(\eta \tilde{W}_t(u) - \frac{1}{2} \mathbb{E}^{\mathbb{P}}[|\eta \tilde{W}_t(u)|^2] \right) = \mathbb{E}^{\mathbb{P}}[V_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t(u) \right),$$

where $\tilde{W}_t(u) := \sqrt{2H} \int_t^u (u-s)^{H-1/2} d\hat{W}_s$ is a square-integrable continuous martingale with $\hat{W}$ a standard Brownian motion under $\mathbb{P}$ such that $\langle \hat{W}, \hat{W} \rangle_t = \rho t$. We also define the stochastic exponential $\mathcal{E} := \exp \left(\Phi - 1/2 \mathbb{E}[|\Phi|^2] \right)$ with Φ a zero-mean Gaussian process. Finally, we define $\eta := 2\nu C_H / \sqrt{2H}$ where $C_H := \sqrt{\frac{2H \Gamma(3/2-H)}{\Gamma(H+1/2) \Gamma(2-2H)}}$. Under an equivalent risk-neutral measure $\mathbb{P}^*$ (with a deterministic change of measure²), Bayer et al. (2016) finally find for the variance process in the rBergomi model

$$\begin{aligned} V_u &= \mathbb{E}^{\mathbb{P}^*}[V_u | \mathcal{F}_t] \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^{1/2-H}} d\hat{W}_s^* - \frac{\eta^2}{2} (u-t)^{2H} \right\}, \\ &= \mathbb{E}^{\mathbb{P}^*}[V_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^*(u) \right), \end{aligned} \quad (1.3.3)$$

where $\hat{W}^*$ is a standard Brownian motion under $\mathbb{P}^*$ and where $\mathbb{E}^{\mathbb{P}^*}[V_u | \mathcal{F}_t]$ is the forward variance curve $\xi_t(u)$ observed on the market. In equation (1.3.3), it is important to note the presence at $t = 0$ of a Riemann-Liouville fBm $\int_0^u C_H (u-s)^{H-1/2} d\hat{W}_s^*$ as defined in Chen et al. (2011), with a Hurst index H that induces the rough behavior of the variance process V . In particular, its paths are $(H-\varepsilon)$ -Hölder continuous ($\varepsilon > 0$), as classical fBm. Moreover, since $\mathbb{E}^{\mathbb{P}^*}[V_u | \mathcal{F}_t] \neq \mathbb{E}^{\mathbb{P}^*}[V_u | V_t]$, this model is non-Markovian with respect to the current variance state V_t . Nevertheless, given the state vector $\xi_t(u)$, which can in principle be computed from observed

² $d\hat{W}_t = d\hat{W}_t^* + \lambda(s)ds$, where $\lambda(s)$ is a deterministic function of s .

option prices (cf. section 1.4.3), the dynamics of the model are well-determined. Finally, the pricing model to be simulated under $\mathbb{P}^*$ from initial time 0 is

$$\begin{aligned} S_t &= S_0 e^{rt} \mathcal{E} \left(\int_0^t \sqrt{V_u} dW_u^* \right), \\ V_t &= \xi_0(t) \mathcal{E} \left(\eta \tilde{W}_0^*(t) \right) = \mathbb{E}^{\mathbb{P}^*} [V_t | \mathcal{F}_0] \mathcal{E} \left(\eta \tilde{W}_0^*(t) \right), \end{aligned} \quad (1.3.4)$$

where $W_t^* = \rho \hat{W}_t^* + \sqrt{1 - \rho^2} W_t^{\perp,*}$ and $W^{\perp,*}$ a Brownian motion under $\mathbb{P}^*$, independent from $\hat{W}^*$. We can hence model the leverage effect in the rBergomi model with $\rho < 0$. It is important to note that the assumption $\lambda = 0$ prevents us from having a stationary model for the volatility and variance processes. The long-term behavior of the rBergomi model will be discussed more thoroughly in Section 1.3.3.

1.3.2 Rough Heston

First, a reminder of the standard Heston model is given in Gatheral (2011) for interested readers. Bergomi and Guyon (2012) show that this Heston model can be written in terms of the forward variance curve $\xi_t(u)$ (see Appendix 1.B.3) by

$$d\xi_t(u) = \nu e^{-\lambda(u-t)} \sqrt{V_t} d\hat{W}_t^*. \quad (1.3.5)$$

Furthermore, from the classical Mandelbrot and Van Ness (1968) representation of the fBm, we clearly see that the kernel $(u-s)^{H-1/2}$ plays a central role in the rough dynamic of the fBm with $H < 1/2$. Indeed, as said above, one can show that the Riemann-Liouville fBm $\int_0^u C_H (u-s)^{H-1/2} d\hat{W}_s^*$ has Hölder regularity $H - \varepsilon$ for any $\varepsilon > 0$. Therefore, in order to allow for a rough behavior of the variance process in a Heston-type model, El Euch and Rosenbaum (2019) naturally introduce the kernel $(u-s)^{H-1/2}$ in the risk-neutral stochastic differential equation of the Heston model with $C_H := 1/\Gamma(H + 1/2)$, leading to the following (generalized) rough Heston model under $\mathbb{P}^*$ for $u \geq t$:

$$V_u = V_t + \frac{\lambda}{\Gamma(H + 1/2)} \int_t^u \frac{\theta^t(s) - V_s}{(u-s)^{1/2-H}} ds + \frac{\nu}{\Gamma(H + 1/2)} \int_t^u \frac{\sqrt{V_s}}{(u-s)^{1/2-H}} d\hat{W}_s^*, \quad (1.3.6)$$

where $\theta^t(\cdot)$ is assumed to be continuous, $\mathcal{F}_t$ -measurable and represents a time-dependent mean reversion level. When $H = 1/2$, one can verify that we indeed recover the classical Heston model. It can also be shown that the trajectories of the volatility itself are almost surely Hölder-continuous of order $H - \varepsilon$, for any $\varepsilon > 0$.

Moreover, El Euch and Rosenbaum (2018) and El Euch et al. (2019) show that $\lambda \theta^t(\cdot)$ can be directly inferred from the forward variance curve observed at time t on the market $\xi_t(u)$, $u \geq t$. By doing so, they can rewrite the model (1.3.6) in the asymptotic setting $\lambda \rightarrow 0$ by

$$V_u = \xi_t(u) + \frac{\nu}{\Gamma(H + 1/2)} \int_t^u (u-s)^{H-1/2} \sqrt{V_s} d\hat{W}_s^*. \quad (1.3.7)$$

Since the forward variance curve $\xi_0(u)$ at time $t = 0$ is observed (or at least can be retrieved) from the market, we only have three parameters left to estimate : H , ν and ρ . Reduction of parameters for the rough Heston is of utmost importance for improving the efficiency of the calibration methodology. Moreover, the fact that we consider $\lambda = 0$ implies that the average long-term behavior of the variance process in the rough Heston will be ruled by the forward variance curve $\xi_0(t)$ and not by the mean-reversion parameters λ and $\theta^0(\cdot)$ anymore (see Section 1.3.3 for more details). We also note that the limit $u \rightarrow t$ of the rough Heston does not make sense. This reflects the fact that this model (as the rBergomi) is not Markovian with respect to the current variance state V_t . However, it is directly visible from equation (1.3.7) that the rough Heston is Markovian in the forward variance curve $\xi_t(u)$. Finally, the rough Heston can also be expressed in the forward variance form as

$$d\xi_t(u) = \frac{\nu}{\Gamma(H + 1/2)} (u - t)^{H-1/2} \sqrt{V_t} d\hat{W}_t^*,$$

which is highly similar to (1.3.5), with the presence of a Riemann-Liouville kernel in addition.

Furthermore, from Gatheral and Keller-Ressel (2019), we have that the Heston and rough Heston forward variance models are said to be affine (see also Chapter 2). In both cases, they show that the characteristic function of the log-asset price $X_T = \log S_T$ at time t can be written as

$$\Phi_t(a, T) = \mathbb{E} \left[e^{iaX_T} | \mathcal{F}_t \right] = \exp \left(ia(X_t + r(T - t)) + \int_t^T \xi_t(s) g(T - s, a) ds \right), \quad (1.3.8)$$

where $g(\cdot, a) : \mathbb{R}_+ \rightarrow \mathbb{R}_-$ is the unique global continuous solution of a convolution Riccati equation, see Gatheral and Keller-Ressel (2019). In the classical Heston model, they find that

$$g(t, a) = \partial_t h(t, a) + \lambda h(t, a),$$

where $h(\cdot, a) : \mathbb{R}_+ \rightarrow \mathbb{R}_-$ is the unique C^1 -function solving the Riccati ODE

$$\partial_t h(t, a) = -\frac{1}{2}a(a + i) - (\lambda - i\rho\nu)h(t, a) + \frac{1}{2}\nu^2 h(t, a)^2, \quad h(0, a) = 0. \quad (1.3.9)$$

In the rough Heston model with $\lambda = 0$, we have this time that $h(\cdot, a)$ is the unique continuous solution of the following fractional Riccati equation

$$g(t, a) = D^\alpha h(t, a) = -\frac{1}{2}a(a + i) + i\rho\nu h(t, a) + \frac{1}{2}\nu^2 h(t, a)^2, \quad I^{1-\alpha} h(0, a) = 0, \quad (1.3.10)$$

with D^α the Riemann-Liouville fractional derivative of order $\alpha = H + 1/2$ and $I^{1-\alpha}$ the Riemann-Liouville fractional integral of order $1 - \alpha$, see Appendix 2.A in Chapter 2. This equation is exactly the same as in the classical Heston model (with

zero-mean reversion) but with the time derivative replaced by a fractional derivative that makes the model rough.

We thus have a quasi-closed formula for the characteristic function in the rough Heston model since it is given by the simple expression (1.3.8). Contrary to the classical Heston case, the drawback is that there is no explicit solution to (1.3.10). However, this fractional Riccati equation can be solved numerically quasi instantaneously using the method described in Gatheral and Radoičić (2019), built upon the combination of a short-time and an asymptotic expansion of the solution $h(t, a)$. As shown in Gatheral and Radoičić (2019), this method is particularly fast, simple and accurate to compute the approximate solution of the fractional Riccati equation (1.3.10). We refer to Section 5 of Gatheral and Radoičić (2019) for a more thorough discussion on the quality and convergence of their approximation for $h(a, t)$. Then, European option prices may be obtained efficiently from the characteristic function (1.3.8) using standard Fourier techniques such as the Carr and Madan (1999) formula. Finally, this fractional Riccati ODE also allows to study portfolio insurance strategies, as developed in the next chapter.

1.3.3 Long-term behavior of rBergomi and rough Heston models

As shown in Chen et al. (2011) (see also Section 1.C.2), the Riemann-Liouville fBm in equation (1.3.4) for the rBergomi and in equation (1.3.7) for the rough Heston is a non-stationary process with non-stationary increments, which thus also makes the variance process of both models non-stationary. Hence, we need to verify whether the variance process exhibits a reasonable long-term behavior in the rBergomi and rough Heston models since we do not have this stationary property anymore as we have in the RFSV model with $\lambda \neq 0$, which ensures reasonable and stable properties on the long-run for the RFSV variance process, see Gatheral et al. (2018). More precisely, we have to make sure that this variance process (and particularly its two first conditional moments) do not explode for large maturities.

First, as shown in Coutin and Pontier (2007) and in Abi Jaber and El Euch (2019), the Riemann-Liouville fBm can be represented as an infinite superposition of Ornstein-Uhlenbeck processes with different mean reversion speeds. Hence, even if $\lambda = 0$, Abi Jaber (2019) shows that there is an inherent mean reversion feature around the forward variance curve in rough volatility models coming from the Riemann-Liouville fBm. This mean-reversion property combined with an appropriate behavior of the long-end of the forward variance curve may thus allow for a sufficiently well behaved variance process at large maturity in rough volatility models, as we will see. For both the rough Heston and the rBergomi models, the fact that we consider $\lambda = 0$ implies that the average long-term behavior of the variance process will be ruled by the forward variance curve $\xi_0(t)$ and not anymore by the mean-reversion parameters λ and $\theta^t(\cdot)$ of equation (1.3.6) for the rough

Heston, nor by the parameters λ and θ of equation (1.3.2) for the rBergomi. Indeed, this simply comes from the fact that the conditional mean of the variance process is now given by $\mathbb{E}^{\mathbb{P}^*}[V_u|\mathcal{F}_t] = \xi_t(u)$. Hence, we will discuss in Section 1.4.3 how to build such forward variance curves with non-exploding long-term behavior. However, long-term behavior of the variance process in both models is not solely driven by the form of this forward variance curve since this curve only controls the average value of the variance process. The non-stationary Riemann-Liouville fBm has also a strong impact on the conditional variance of the variance process given by $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$, as explained in Appendix 1.C.2. Indeed, this quantity $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ increases exponentially with u in the rBergomi model and hence tends to exhibit extreme values for large maturities (see Figure 1.C.1). For the rough Heston model, the rate of increase of this quantity is much slower and hence, we do not observe this exponential growth of $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ for the considered maturities, as again depicted on Figure 1.C.1. The generated variance sample paths at large time scales are thus more reasonable and less volatile in the rough Heston while they exhibit extremely large swings for the rBergomi. This behavior will have an impact on the price of long-term life insurance claims, as we will see in the last section of this chapter.

1.4 Calibration

In this section, we will show how to build a consistent calibration methodology. We will then compare four models (Heston, Bates, rBergomi and rough Heston) in terms of fit of the observed European option prices and implied volatility surface for the CAC 40 index. We will mainly focus on the long-term behavior of such models and we will then use these calibrated models to price our equity-linked endowment embedding the cliquet option (1.2.1).

As in Gatheral et al. (2018), we first provide a statistical estimation of the RFSV parameters based on historical data for the CAC 40 index from January 2004 to July 2020 (obtained from the Oxford-Man Institute). Using the scaling property of Gatheral et al. (2018) to estimate the smoothness of the volatility process for the CAC 40, we find an estimated value $\hat{H} = 0.13$. The historical estimate of ν is also derived from this scaling property and is equal to $\hat{\nu} = 0.31$.

The aim of the calibration method is to find, for each model, the model parameters that match derivative model prices as best as possible with the observed market prices of vanilla derivatives in the market. It then boils down to an optimization problem where we have to find the minimum distance between model prices/model-based implied volatility and market prices/market implied volatility. In this first chapter, we choose to minimize the distance between model-based implied volatility $\hat{\sigma}_j^{imp}$ and market implied volatility σ_j^{imp} of the j^{th} derivative ($j = 1, \dots, N$) since it leads to more stable results. Finally, we introduce the weighted Root Mean

Square Error as the loss function, which is defined as

$$\text{RMSE} = \sqrt{\sum_{j=1}^N \frac{w_j}{N} \left(\sigma_j^{\text{imp}} - \hat{\sigma}_j^{\text{imp}} \right)^2}, \quad (1.4.1)$$

where $w_j = \frac{1}{\text{Ask}_j - \text{Bid}_j}$, *i.e.* the weight is the inverse of the bid-ask spread, expressed in terms of implied volatility. This way, we give more importance to observations with small bid-ask spread. Indeed, if an option is liquid, then its bid-ask spread is supposed to be small and its price/implied volatility is more accurate. Mathematically, it boils down to find the optimal set $\Theta^* \in \mathbb{R}^p$ of the p model parameters such that

$$\Theta^* = \arg \min_{\Theta} \mathcal{L} \left(\sigma^{\text{imp}}, \hat{\sigma}^{\text{imp}}(\Theta) \right),$$

where $\mathcal{L}$ is the RMSE loss function. Finally, we highlight the fact that the results of the optimization methodology can be highly dependent on the initial parameters. Even for the Heston model, there is no consensus among researchers on whether the objective function for the calibration is convex or not, as explained in Cui et al. (2017). In any case, to overcome this problem of guessing the initial parameters, we use a global optimizer³ combined with a local optimizer (the standard *nlminb* function in R). Indeed, global optimizers can find a solution on their own even without any initial guess while local optimizer need a suitable starting point. Therefore, we use the approach described in Mrázek et al. (2016) which takes the final parameters provided by the global optimizer as initial values for the local optimization. This combination of global and local optimizers allows to refine the final parameters of the calibration and minimize the loss function. However, even with this technique, we will see that the rBergomi model still appears to be highly unstable due to the need of Monte-Carlo simulations for this model.

1.4.1 Market data

We first choose as starting date $t = 0$, the 6th of July 2020. The maturity of vanilla options as well as the risk-free term structure will be defined with respect to this date.

Firstly, the risk-free rates used for pricing purposes are derived from the observed swap term structure. The technique for constructing this swap term structure divides the curve into two term buckets. The short end of the swap term structure is built using interbank deposit rates. We will here consider EURIBOR rates from one day to 12 months. The long end of the risk-free interest rate curve (maturity above 1 year) will be based on the European swap term structure available from

³More precisely here, the Simulated Annealing (SA) algorithm of Van Laarhoven and Aarts (1987) with the package *GenSA* in R.

the EIOPA for the month of July⁴. Table 1.C.1 in Appendix provides a summary of the risk-free rates used between 0 and 20 years. The dividend yield is chosen to be 1% but this assumption does not have a significant impact on the results and conclusions of this chapter.

Secondly, we have that the price S_0 at time 0 of the CAC40 index is equal to 5028.56 €. As mentioned above, we also need market prices of European options on the CAC40 in order to calibrate later our model since these are the most traded financial instrument in the equity world. Indeed, there are no liquid long-term products traded on the market on which we can calibrate our models for large maturities. The best we can hope so as to be the most market consistent is therefore to reproduce as best as possible the observed implied volatility surface (and hence the prices of European options) and then extrapolate the obtained trend for large maturities. In particular, we will see two methods below in Section 1.4.3 for obtaining the values of the forward variance curve beyond the available maturities. From Bloomberg, we thus gather a database that includes 366 European options with the option bid price and the option ask price for different strikes and maturities. The true market price is here considered as the mid value between the bid price and the ask price, which is a common assumption. Using the one-to-one relationship between market prices and implied volatility from the Black and Scholes formula, we can build a database with the bid, ask and mid implied volatility for each strike and maturity. We then apply the following common filters as in Moyaert and Petitjean (2011) and select the following options :

- **Out-of-the-money options** : Since out-of-the-money options are more actively traded than in-the-money options (as a protection), the quotes on out-of-the-money options are usually more reliable.
- **The bid-ask spread** (on the option prices) is less than 5%. Otherwise, we consider that there is no sufficient liquidity to take the option into account.
- **Maturity** : We reject options with maturity equal to 0.9 year due to data quality issue for this particular maturity.

We have a final database of 145 OTM options for which we can observe the strike, the maturity, the bid, ask and mid implied volatility. We now turn to the specific calibration method for each model. Indeed, even if the general methodology is the same, the way we price our European options will vary depending on the model.

1.4.2 Heston and Bates models

We have at disposal for these two models a closed-form expression for their characteristic function, as reviewed in Gatheral (2011). Then, for a given maturity,

⁴Further information on risk-free rates computation can be found at : https://www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en

the Carr-Madan formula generates prices of European options for a whole range of strikes in one run very fast using Fourier inversion. By calling many times this pricing algorithm on the available maturities in the data, we can thus find the optimal parameters of each model.

- **Heston :**

A reminder of the standard Heston model is given in Gatheral (2011). The set of initial parameters obtained from the Simulated Annealing algorithm is given by (0.2898, 2.9176, 0.0966, 1.4703, -0.7010). Using the local optimizer *nlm* in R, we then find a minimal RMSE of 0.11694 with the following final parameter values Θ^* :

σ_0^*	λ^*	θ^*	ν^*	ρ^*
0.28992	2.91760	0.09664	1.47027	-0.7010

We see that we have a high speed of mean reversion λ^* and a high level of the vol-of-var parameter ν^* which implies large swings in the variance process. We also obtain a large negative correlation between the stock process and the variance process, which is consistent with the leverage effect. Finally, the mean reversion level θ^* towards which the variance will converge is 9.66% (which represents a volatility of 31,09%). It is no wonder to obtain this high level of volatility as well as vol-of-var parameter given the state of the economy due to the COVID-19 outbreak in July 2020. Finally, the Heston model with the parameters obtained above leads to Figure 1.1 when compared to market prices of European options on the CAC 40.

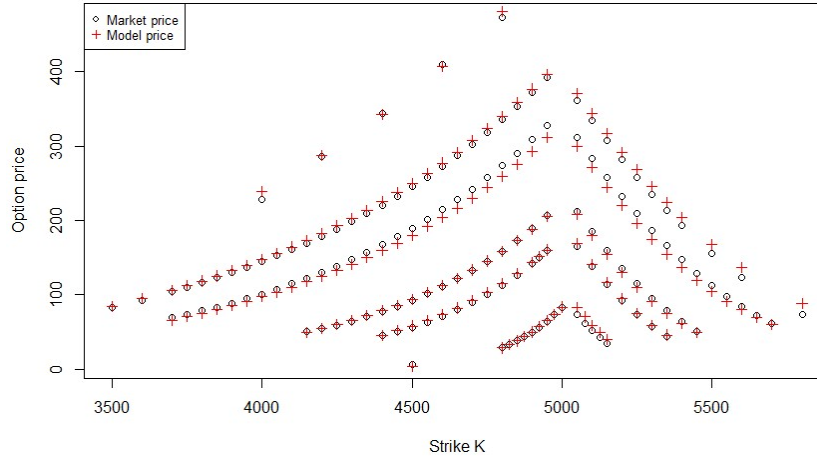


Figure 1.1: Comparison of Heston option prices in red crosses vs. market option prices in black dots (CAC 40).

The calibration of the forward-variance specification of the Heston model, as de-

scribed in Bergomi and Guyon (2012) and Buehler (2006), is provided in Appendix 1.B.3. This Appendix will allow to compare more thoroughly the Heston model with rough volatility models which are also based on this forward variance curve.

- **Bates :**

For a reminder of the Bates (or SVJ) model, we refer to Gatheral (2011). The set of initial parameters obtained from the SA global optimizer is given by (0.27, 0.62, 0.12, 1.24, -0.66, -0.10, 0.11, 0.82). Using our local optimizer, we then find a minimal RMSE of 0.08465 with the following final parameter values Θ^* :

σ_0^*	λ^*	θ^*	ν^*	ρ^*	μ_j^*	σ_j^*	λ_j^*
0.2589	0.1555	0.2709	1.1593	-0.6640	-0.1004	0.1072	0.9080

Firstly, the RMSE is substantially lower than the one obtained with the Heston model, which means that the Bates model provides a better fit of the observed European option prices. The mean-reversion speed λ^* and the volatility of variance parameters ν^* are also lower, which implies that the swings in the variance process are less pronounced than in the Heston model. Moreover, we now observe jumps in the price process that are on average negative since $e^{\mu_j + \sigma_j^2/2} - 1 < 0$. The annual frequency of these jumps is equal to 90.8%. Finally, we observe that the long-run variance θ^* is higher than in the Heston model. This Bates models leads to Figure 1.2 when compared to the observed market prices. The fit is clearly better than

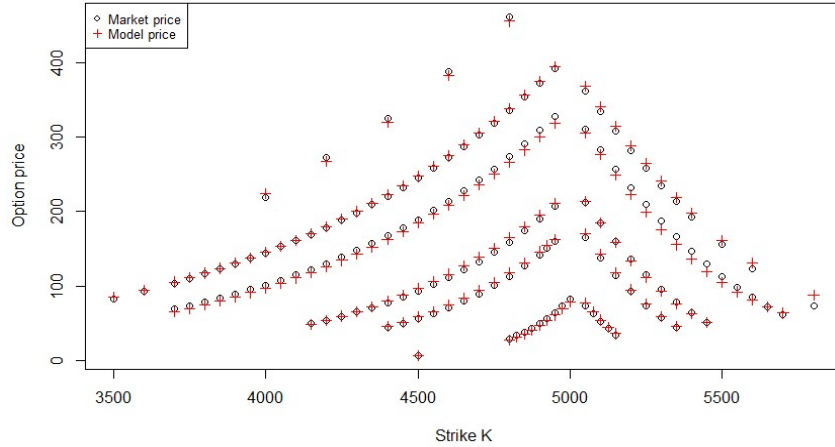


Figure 1.2: Comparison of Bates option prices in red crosses vs. market option prices in black dots (CAC 40).

in the Heston model as said above. More precisely, we see that this improvement comes from the better fit of OTM call options with large maturities when using the Bates model. However, we also observe on Figure 1.2 that option prices with the

third largest maturity ($T = 0.45$) are still significantly out-of-market. Note that the presence of jumps makes the calibration of the forward-variance specification of the Bates model really difficult and time-consuming. Indeed, the characteristic function of the Bates model cannot be written as an integral of $\xi_t(s)$ as in the Heston and rough Heston case with equation (1.3.8) since the Bates model is not an affine forward variance model due to its jump term, see Gatheral and Keller-Ressel (2019). We will thus not consider the forward-variance specification of the Bates model in this chapter.

1.4.3 Rough Heston calibration

As explained above, we have a quasi-closed form for the characteristic function $\Phi_t(a, T)$ of the log-price X_T under the rough Heston model (1.3.7). We can then apply exactly the same calibration technique as for the Heston and Bates models described in the previous section. However, recall that the rough Heston is not Markovian in the current variance state V_t but Markovian in the (forward) variance swap curve $\xi_t(u)$. Hence, the rough Heston's characteristic function depends on this variance swap curve. In practice however, the variance swap curve $\xi_0(u)$ at time $t = 0$ is not obtained directly from the financial markets. Indeed, it is hard to obtain high-quality variance swap data since variance swaps are OTC contracts. We thus choose to proxy the value of a variance swap of maturity T by the value of a log-contract (also of maturity T), as described *e.g.* in Chapter 11 of Gatheral (2011). Therefore, even if variance swaps are OTC contracts, we can obtain variance swap prices directly from the observed European options which are heavily traded on the market. This way, we can consider in this context that the variance swap quotes derived from our European options are reliable. The first method to obtain the initial variance swap curve is therefore to calibrate it based on our dataset of observed option prices for the CAC 40. We hence obtain in Figure 1.3 the following market variance swap curve, denoted $\xi_0^{mkt}(t)$, between 0 and 1.2 years when interpolated between the observed maturities in black circles.

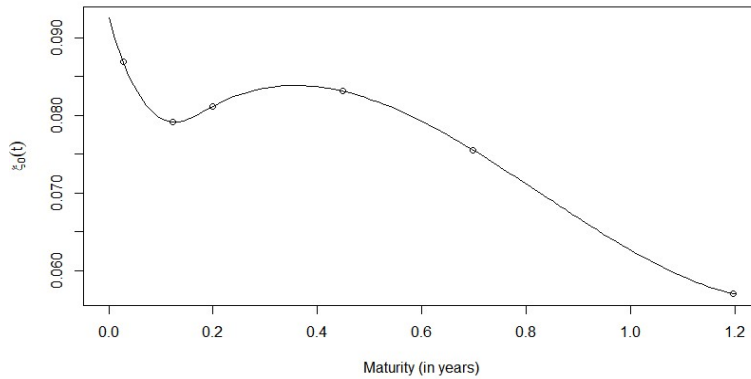


Figure 1.3: Initial market forward variance curve $\xi_0^{mkt}(t)$ between 0 and 1.2 years.

However, we only have European options with observed maturities until 1.2 years. In order to price long-term claims on the underlying, we need to extrapolate the values of $\xi_0^{mkt}(t)$ beyond 1.2 years and hence, we need to consider a term-structure parametrization of $\xi_0^{mkt}(t)$. Since the forward variance curve is estimated from the data and is not an output of the model, the average long-time behavior of the variance process is embedded in this forward variance curve parametrization (see Section 1.3.3). Indeed, recall that the forward variance curve controls the average value of the variance process V since $\xi_0(t) := \mathbb{E}^{\mathbb{P}^*}[V_t|\mathcal{F}_0]$. The long-end of this curve will thus have a strong impact on the behavior of the model for large maturities and we hence need to avoid parametrizations generating explosive or unrealistic forward variance curves. Otherwise, the variance process (and hence also the price process) will tend to explode when the maturity increases. The easiest way to achieve this is to impose a constant long-term value for the initial forward variance curve $\xi_0^{mkt}(t)$. Therefore, it is reasonable to choose a parametrization with asymptotic line such that

$$\lim_{t \rightarrow \infty} \xi_0^{mkt}(t) = cst.$$

This assumption allows the variance process of the rough Heston to be in a stable regime for large maturities. A classical choice for such parametrization is the Gompertz function

$$\xi_0^{mkt}(t) = z_1 e^{-z_2 e^{-z_3 t}}, \quad (1.4.2)$$

where z_1 is the asymptote, z_2 sets the displacement along the x-axis (here, the time to maturity) and z_3 sets the growth rate. Fitting the Gompertz function to the observed forward variance curve, we find :

$$z_1 = 0.03954 \quad z_2 = -0.7992 \quad z_3 = 0.5672$$

Therefore, the asymptotic level of the volatility is given by $\sqrt{0.03954} = 19.88\%$. Finally, combining the curve obtained in Figure 1.3 with the Gompertz fit for $t > 1.2$, we obtain the initial market forward variance curve $\xi_0^{mkt}(t)$ in Figure 1.4. We can observe that we enter a stable regime as of the 6th year.

Yet, we are aware that calibrating such Gompertz function based on only 6 different maturity points can seem dubious. Indeed, estimating an asymptotic value for the forward variance from the information carried out by short-maturity option data does not provide a lot of confidence in our results. Moreover, this way of building and extrapolating the forward variance curve does not ensure to avoid static arbitrage. Therefore, we now show how to tackle both of these issues. A better alternative to estimate the forward variance curve is to fit the SSVI parametrization from Gatheral and Jacquier (2014) of the observed implied volatility surface and compute the forward variance curve associated with this parametrization. A brief reminder of such arbitrage-free SSVI parametrization is given in Appendix 1.A. The advantage of this procedure is to ensure that the estimated forward variance curve is consistent with the absence of calendar spread and butterfly arbitrages.

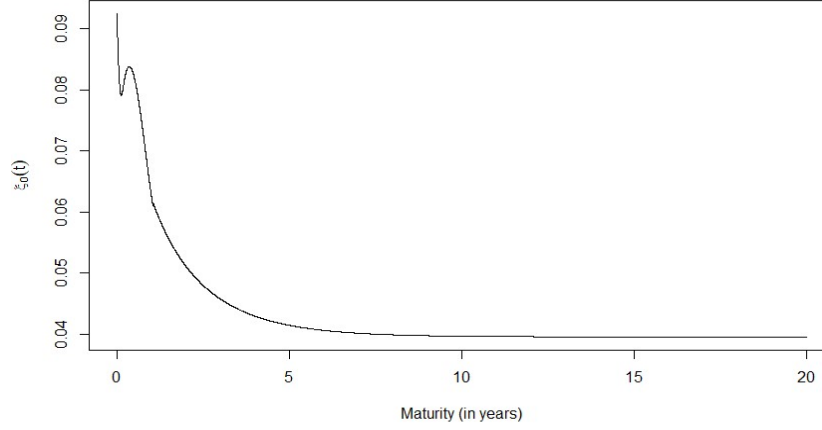


Figure 1.4: Initial market forward variance curve $\xi_0^{mkt}(t)$ until $t = 20$ years.

Furthermore, the extrapolation of the forward variance curve does not require to specify a particular function for this curve anymore since its extrapolation beyond 1.2 years in the SSVI parametrization only relies on an arbitrage-free argument, as explained again in Appendix 1.A. We therefore do not need anymore to determine a long-term value of $\xi_0(t)$ based solely on short-term maturity data. We obtain this way the following SSVI forward variance curve $\xi_0^{SSVI}(t)$ depicted on Figure 1.5. We

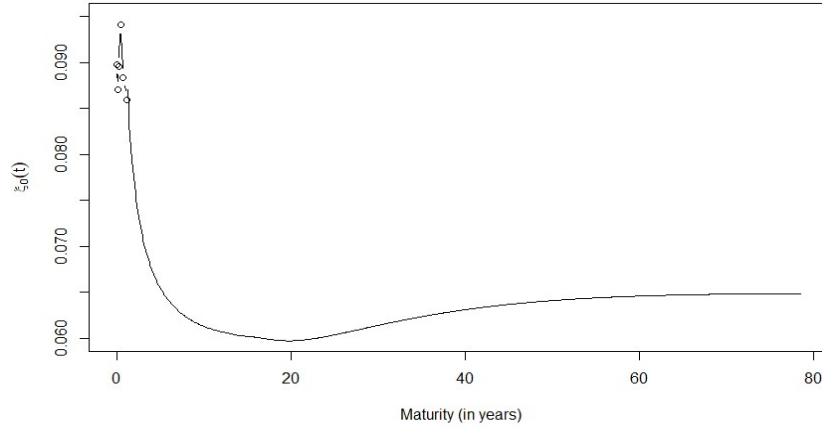


Figure 1.5: Initial SSVI forward variance curve $\xi_0^{SSVI}(t)$ until $t = 80$ years. Black circles are the variance swap values for the available maturities.

see that this forward variance curve $\xi_0^{SSVI}(t)$ tends to approximately exhibit the same shape as the market forward variance curve between 0 and 80 years. More-

over, the SSVI forward variance curve appears to be stable and does not explode for large maturities, which also ensures a stable regime and a reasonable average behavior of the rough Heston model in the long term.

The last alternative we consider is to use a constant forward variance curve denoted $\xi_0^{cst}(t)$. We decide to fix arbitrarily $\xi_0^{cst}(t) = \xi_0^{mkt}(0) = 0.0925$, $\forall t \in [0, T]$. Note that this constant level is very close to the long-term mean θ^* obtained in the Heston calibration ($= 0.0966$). This parametrization for the forward variance curve will hence allow for a more thorough comparison of rough volatility models with the latter. Calibration and pricing results with $\xi_0^{cst}(t)$ are given in Appendix 1.B.1.

The SSVI forward variance curve being arbitrage-free, more reliable and more consistent, we now perform the calibration of the rough Heston model based on $\xi_0^{SSVI}(t)$. Results for the market and the constant forward variance curves are given in Appendix 1.B. Using the Carr-Madan formula and the numerical approximation of Gatheral and Radoičić (2019) for the fractional Riccati equation (1.3.10), we obtain in a first step the following set of initial parameters from the SA global optimizer : (0.1702, 0.6241, -0.6724). Using our local optimizer, we then find a minimal RMSE of 0.08457 with the following (same) final parameter values Θ^* :

H^*	ν^*	ρ^*
0.1703	0.6241	-0.6725

Based on the RMSE, we then have a slightly better fit than with the Bates model. It is impressive knowing that the rough Heston model in the asymptotic setting $\lambda \rightarrow 0$ has only 3 effective parameters to estimate while the Bates model has 8 parameters ! We will also analyze in Section 1.5 which of the models best fits the behavior of the implied volatility surface (and particularly the term structure of ATM skews). Moreover, it is important to note that the Hurst index H^* obtained via the risk-neutral calibration method is consistent with the Hurst index obtained via statistical estimation (under the real-world measure) equal to 0.13. Yet, we have a higher ν^* than the one estimated based on the historical time series ($= 0.31$) due to the market price of risk included in the risk-neutral valuation. Finally, we obtain a leverage effect of the same order as in the Heston and Bates models. Figure 1.6 depicts the fit of the rough Heston options prices compared with the observed market prices of European options.

1.4.4 rBergomi Calibration

The calibration of this model departs significantly from the three models described above. Indeed, no characteristic function is available in (quasi-)closed form. We then rely on Monte-Carlo simulations with the Hybrid scheme of Bennedsen et al. (2017), as applied in McCrickerd and Pakkanen (2018), to calibrate the parameters ρ , η and H of the rBergomi model under $\mathbb{P}^*$. More precisely, since the Riemann-Liouville fBm is a special case of Brownian Semistationary (BSS) processes, we

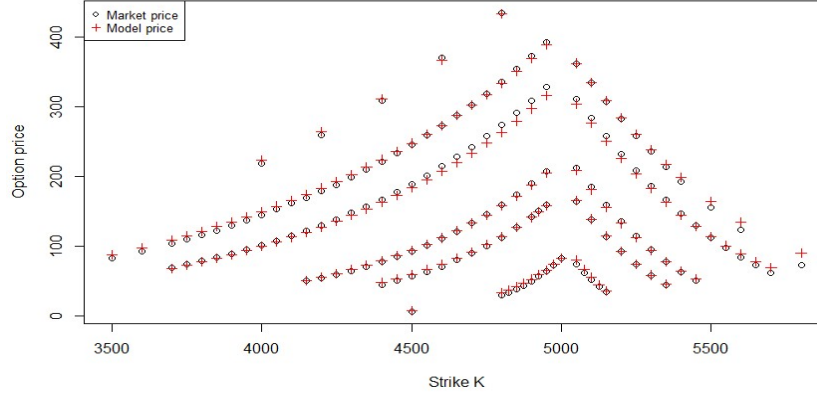


Figure 1.6: Comparison of rough Heston option prices vs. market option prices (CAC 40)

follow the papers of Bennedsen et al. (2017) and of McCrickerd and Pakkanen (2018) in order to improve the efficiency of simulations in the rBergomi model. Note that the initial forward variance curve is exactly the same as the one derived in the calibration of the rough Heston model. Hence, we again use $\xi_0^{SSI}(t)$ in this section.

We find a RMSE of 0.08631 with the initial set of parameters equal to (0.15, 1.67, -0.92). This leads to the following final values Θ^* :

H^*	η^*	ρ^*
0.1515	1.6803	-0.9275

First, the RMSE is slightly higher compared with the rough Heston and Bates models but lower than the Heston model. We see that the calibrated Hurst exponent H^* is lower than the one found in the rough Heston model and closer to the historical estimate ($= 0.13$). Based on a $\eta^* = 1.6803$, we find the corresponding $\nu^* = \eta^* \sqrt{2H^*} / (2C_H) = 0.9969$. Therefore, the volatility of variance parameter ν^* clearly needs to be much higher than the historical $\hat{\nu}$ to better fit the European option data (again due to the volatility of variance risk premium). The anti-correlation between the stock process and the variance process is stronger than the ρ^* derived from the rough Heston. Moreover, it is important to add that these results are quite unstable from simulation to simulation. The hybrid scheme is indeed not yet fast enough to provide a reliable calibration method, at least in our implementation. We used $m = 1,000,000$ and $\Delta t = t_k - t_{k-1} = T/n$ with $n = 400$ and the optimization was done in more than 5 hours with these parameters. We need to increase m and decrease Δt to obtain more precise and stable results but it then requires far more computing time and memory, which is not suitable and tractable in practice. We then obtain Figure 1.7 where we can see that the OTM

puts are substantially off-market with this model. Finally, recall that the SSVI

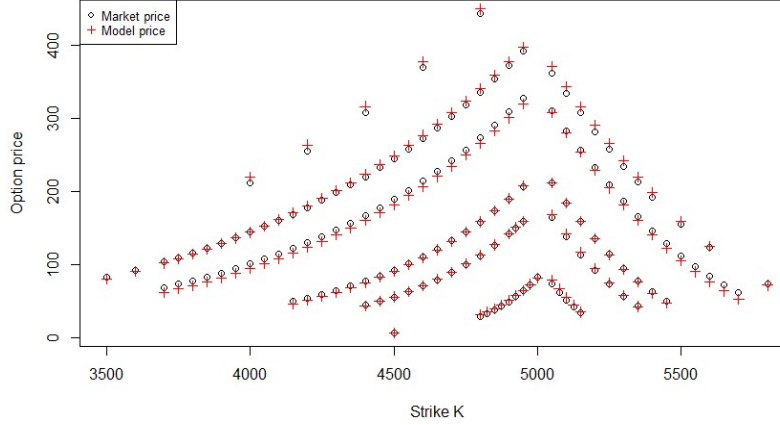


Figure 1.7: Comparison of rBergomi option prices vs. market option prices (CAC 40)

parametrization of the forward variance curve ensures that in the long term we have on average a stable and non-exploding variance process in the rBergomi and rough Heston models, as shown in Figure 1.5. Nevertheless, the presence of the non-stationary Riemann-Liouville fBm in both models also impacts the long-term behavior of the variance process. More precisely, if we now compare the conditional variance of the variance process $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$, we obtain an increasing quantity in u with divergent properties between the two models as shown and explained with Figure 1.C.1 in Appendix 1.C.2 and with the corresponding equations (1.C.1) and (1.C.2). Clearly, the exponential growth of $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ in the rBergomi model does not allow for a reasonable behavior of the variance process in the long term with the calibrated parameters above. The same extreme behavior is obtained using the parameters of the rBergomi model found in the initial paper of Bayer et al. (2016) and in Bayer et al. (2019, b). On the contrary, the conditional variance $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ in the rough Heston model exhibits an almost constant behavior for large times to maturity and is hence more appropriate for long-term pricing. This stable regime of the rough Heston model is also obtained using standard parameters such as in El Euch et al. (2019). The impact of the quantity $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ on the price of life insurance claims for large maturities u will be analyzed more thoroughly in the last section.

1.5 Volatility surface fit

We now compare the Heston, Bates, rough Heston and rBergomi models in terms of implied volatility surface fit. More precisely, we will look at the fit of the term structure of ATM volatility skew $\psi(\tau)$ and confirm that rough volatility models provide a better fit.

From our market option prices database, we can build the term structure of ATM volatility skew $\psi(\tau)$. First, recall that $\psi(\tau) := \left| \frac{\partial}{\partial k} \hat{\sigma}_t^{imp}(\tau, k) \right|_{k=0}$ with k being the log-strike and $\tau = T - t$ the time to maturity. We can approximate it by

$$\hat{\psi}(\tau) = \left| \frac{\hat{\sigma}_0^{imp}(\tau, k + \delta) - \hat{\sigma}_0^{imp}(\tau, k - \delta)}{2\delta} \right|_{k=0},$$

for small enough δ and for each available time to maturity τ . We then obtain in Figure 1.8 the term structure of ATM volatility skew for our data at time $t = 0$. The red dots are the estimated ATM skews for each of the available maturities.

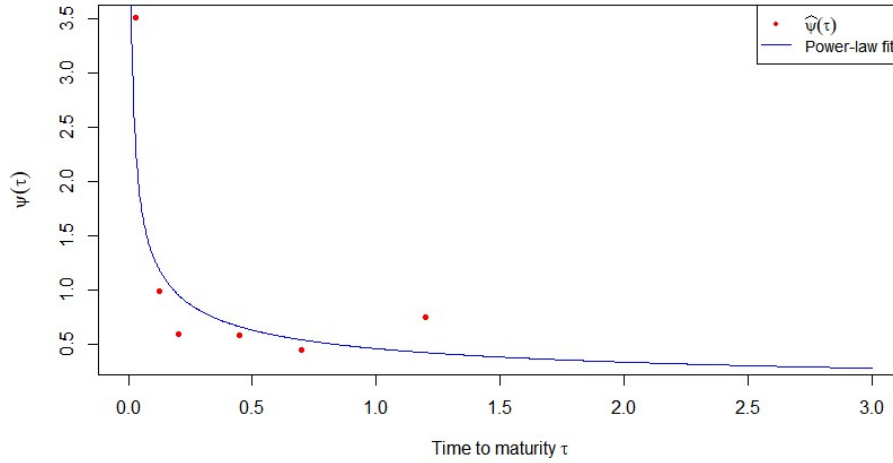


Figure 1.8: Red dots are the ATM volatility skews $\hat{\psi}(\tau)$ estimated for the available maturities. The power-law fit is superimposed in blue

We clearly see that we obtain a term structure of ATM volatility skew which is consistent with the power-law function. Indeed, the blue line is the power-law fit to the data, obtained by the following linear regression

$$\log \psi(\tau) = -\alpha \log \tau.$$

We find $\hat{\alpha} = 0.4546$. From Section 1.3, we know that $\psi(\tau) \sim \tau^{H-1/2}$ holds for the RFSV model. Hence, we have that $\hat{H}_{ATM} = 1/2 - \hat{\alpha} = 0.0452$, which is lower than the Hurst exponent estimated with historical data (but again rough). However, we do not have a lot of different maturities in our option data set, which prevents us from having a consistent estimator $\hat{H}_{ATM}$ using this power-law fit. Therefore, we will mainly focus our analysis on the shape of the ATM skew term structure and on whether our different models are able to approximate this blue line rather than on the exact fit.

1.5.1 Comparison of the fit

We now compare the empirical term structure of ATM volatility skews with the term structures derived from each model. We first analyze the fit of the standard Heston and Bates model thanks to Figure 1.9. The blue line is again the power-law fit of the observed ATM skews (with $\hat{\alpha} = 0.4546$). The green dots are the ATM skews $\hat{\psi}(\tau)$ derived from the different models for the available maturities. For the Heston model, we clearly observe that it does not allow to capture the high values of ATM skews for short maturities and hence, it cannot reproduce the power-law shape of $\hat{\psi}(\tau)$. This was indeed one of the issues of this model that we rose in Section 1.3. For $\tau > 0.20$, we can however note that the Heston fit approximates quite well the empirical blue line. Concerning the Bates model, the presence of jumps only provides a slightly better fit of the ATM skews $\hat{\psi}(\tau)$ compared with the Heston model. Indeed, the Bates model is still not able to capture the explosion of the ATM skews for very short time to maturity. This is due to the fact that the parameter λ_j controlling the frequency of jumps is only of 0.91 and that the average size of the jumps is also rather low. A higher value of λ_j or of $|\mu_j|$ would allow to better capture this explosion phenomenon for $\tau \rightarrow 0$.

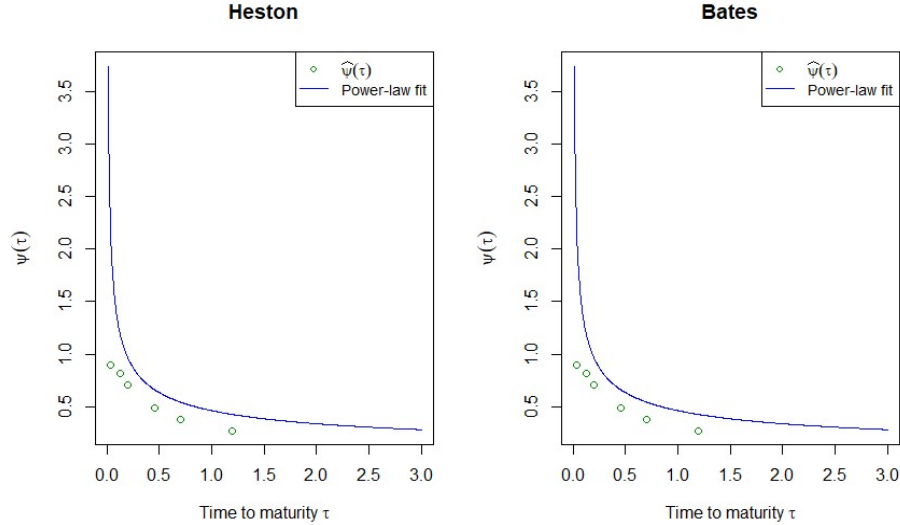


Figure 1.9: Empirical fit (blue line) vs. Heston and Bates fit (green dots) of $\psi(\tau)$, CAC 40 index.

We now analyze the fit of the rough Heston and the rBergomi models since we said that rough volatility models should be able to better capture the high curvature of $\hat{\psi}(\tau)$ for small time to maturity thanks to the rough behavior of their volatility process. We obtain the following Figure 1.10. We clearly see that the green dots

of both figures (*i.e.* the rough Heston and rBergomi fit for available maturities) exhibit the same shape as the blue line and can therefore be fitted by a power-law function of the form $\tau^{H-1/2}$. If we had at disposal even shorter maturities, the rough Heston and rBergomi models would likely better capture the explosion of $\hat{\psi}(\tau)$ for $\tau \rightarrow 0$. We can finally conclude that rough volatility models are more consistent with the observed term structure of ATM volatility skew than standard SV models.

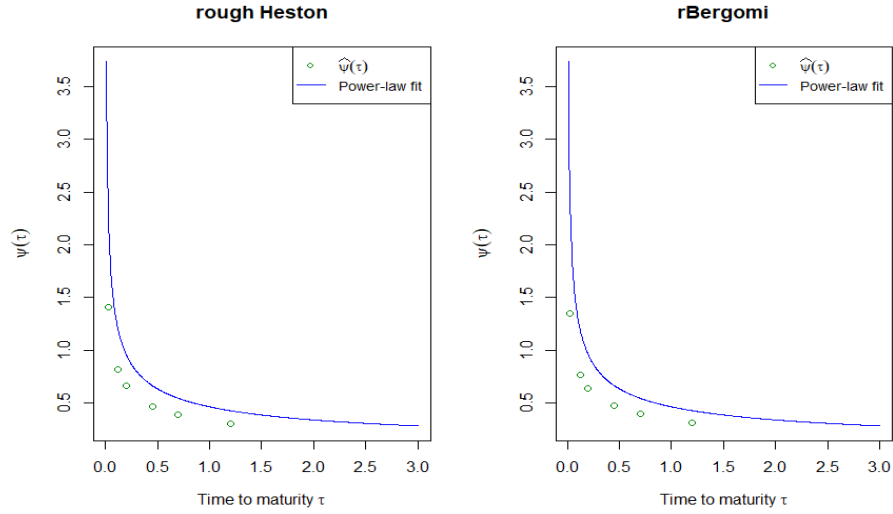


Figure 1.10: Empirical fit (blue line) vs. rough Heston and rBergomi fit (green dots) of $\psi(\tau)$, CAC 40 index.

1.6 Simulation and discretization methodology

In order to price the above equity-linked life insurance contract, we first need to explain how we can discretize and simulate sample paths of the price process and the variance process under our different models. These simulation schemes will also be useful in Chapter 2.

- Heston :

Once the Heston model has been calibrated and the optimal parameters $\sigma_0, \lambda, \theta, \nu$ and ρ have been found, we can use a Monte Carlo approach to simulate the sample paths of the Heston model under $\mathbb{P}^*$. We first simulate the stock price process and the variance process by generating correlated $\mathcal{N}(0, 1)$ random numbers ε_S and ε_V

(with correlation ρ). The relationship between ε_S and ε_V can be written as

$$\varepsilon_{V,t_i} = \rho \varepsilon_{S,t_i} + \sqrt{1 - \rho^2} \varepsilon'_{t_i},$$

where ε_{S,t_i} and ε'_{t_i} are independent $\mathcal{N}(0, 1)$ random variables. Using an Euler discretization scheme for the price process and the variance process, we have

$$S_{t_{i+1}} = S_{t_i} \left(1 + r\Delta t + \sqrt{V_{t_i}} \sqrt{\Delta t} \varepsilon_{S,t_i} \right), \quad (1.6.1)$$

$$V_{t_{i+1}} = V_{t_i} + \lambda(\eta - V_{t_i})\Delta t + \nu \sqrt{V_{t_i}} \sqrt{\Delta t} \varepsilon_{V,t_i}, \quad (1.6.2)$$

for $i = 0, \dots, n-1$, with $t_0 = 0$ and $t_n = T$, where n is the number of time steps and $\Delta t = T/n$, for a given maturity T . We then simulate m sample paths of this stock price process and of this variance process. Now that we have the full path of $(S_{t_i})_{i=0, \dots, n}$ for each m , we can price path-dependent options, such as the cliquet option (1.2.1). In practice, the variance can become negative under a simulation because of the discretization. We then use the full truncation scheme of Lord et al. (2010) to tackle this problem since it overcomes other Euler fixes (absorption, reflection, etc.) and quasi-second order schemes (Milstein scheme, Ninomiya and Victoir (2008) scheme, etc.) in terms of positive bias. The full truncation scheme therefore better reproduces the true properties of the CIR process. This simulation methodology gives us the following Figure 1.11, depicting one sample path of the variance process under the standard Heston model over 10 years. We can clearly see the effect of the high volatility of variance (vol-of-var) parameter ν with large swings in the variance process.

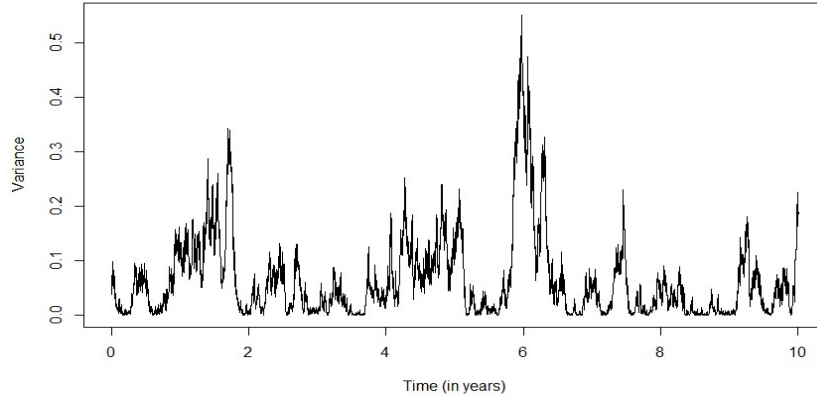


Figure 1.11: One sample path of the variance process under the standard Heston model

- Bates :

We again use the Euler scheme with a full truncation fix for the variance process

as in the Heston model. However, the discretization scheme of the price process is replaced by

$$S_{t_{i+1}} = S_{t_i} \left(1 + r\Delta t + \sqrt{V_{t_i}} \sqrt{\Delta t} \varepsilon_{S,t_i} + (Y_i - 1) \Delta N_{t_i} \right), \quad (1.6.3)$$

where $Y_i \sim \text{Log}\mathcal{N}(\mu_j, \sigma_j^2)$ and $\Delta N_{t_i} \sim \text{Poi}(\lambda\Delta t)$.

- Rough Heston :

The discretized price process is exactly the same as in the Heston model with (1.6.1). However, due to the non-Markovian nature of the variance process under the rough Heston, an Euler discretization scheme for simulating V_t is very slow and does not allow to use the full truncation fix in order to deal with negative variance. A more time-efficient simulation scheme for the rough Heston can be found instead in Abi Jaber (2019) with a lifted version of the Heston model. This lifted Heston model appears to be a multi-factor approximation of the rough Heston model built as an infinite superposition of square-root (CIR) processes with the same dynamic but mean reverting at different speeds (cf. Section 1.3.3). Such infinite-dimensional Markovian representation of the limiting rough variance process provides an appealing trade-off between flexibility and tractability, as explained in Abi Jaber (2019). In practice, only few factors are sufficient which drastically speeds up the simulation procedure compared to an Euler discretization scheme. Moreover, this representation in terms of square-root processes allows to use the full truncation fix of Lord et al. (2010) described above in order to avoid negative variance and to better reproduce the true properties of the variance process. For simulating our rough Heston model, we take the suggested values of the lifted Heston model defined in Abi Jaber (2019): $\tilde{n} = 20$ factors and $r_{20} = 2.5$. The function $g_0^{\tilde{n}}(t)$ of this lifted Heston model is chosen so as to match the forward variance curve $\xi_0(t)$ and the speed of mean-reversion λ is again set to 0 since mean reversion at different speeds is inherent in the lifted Heston model. This parametrization allows to recover the rough Heston model (1.3.7) with its 3 effective parameters. One sample path of the variance process over 10 years under our calibrated rough Heston model with SSVI forward variance curve is plotted in Figure 1.12. We clearly see the rough behavior of the variance process compared to the sample paths of the Bates and Heston models. We also have that the vol-of-var parameter ν is quite high, leading to large movements in the variance process.

- rBergomi :

The price process is again given by (1.6.1). However, the variance process is now simply given by

$$V_{t_i} = \xi_0(t_i) \mathcal{E} \left(\eta \tilde{W}_0^*(t_i) \right),$$

where we use the SSVI forward variance curve for $\xi_0(t_i)$ ($= \mathbb{E}^{\mathbb{P}^*}[V_{t_i}]$) and the hybrid scheme of McCrickerd and Pakkanen (2018) and Bennedsen et al. (2017) for the

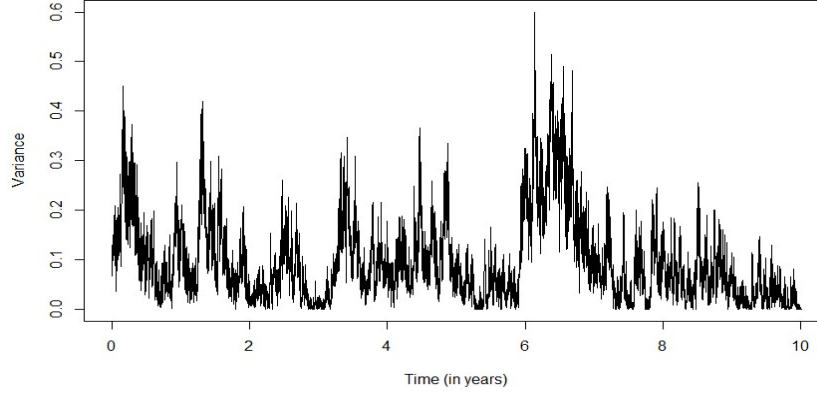


Figure 1.12: One sample path of the variance process under the rough Heston model

simulation of $\tilde{W}_0^*(t_i)$, $i = 0, \dots, n$. The way the variance process is defined in (1.3.4) under the rBergomi model prevents us by definition from having negative variance. You can also find in Figure 1.13, one simulated sample path of the rBergomi variance process using this Hybrid scheme.

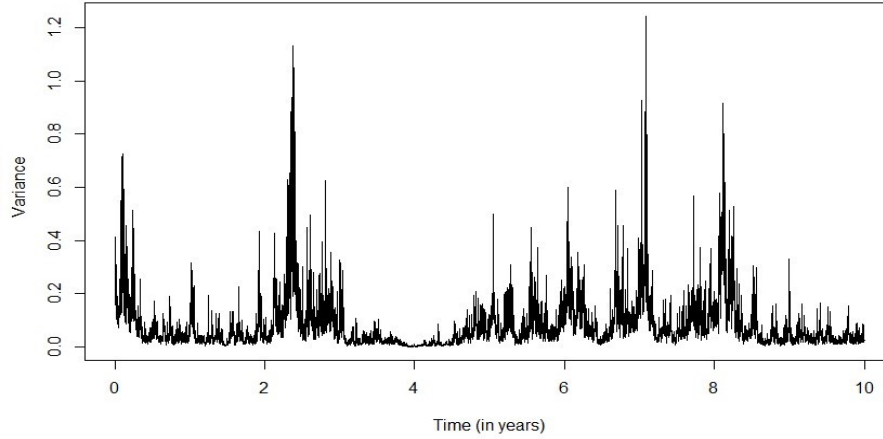


Figure 1.13: One sample path of the variance process under the rBergomi model

Due to the huge vol-of-var parameter η , the variance paths tend to explode and exceed 100%. The generated sample paths with the rBergomi are therefore not realistic. Furthermore, if we compute the sample variance corresponding to $\mathbb{V}^{\mathbb{P}^*}[V_t]$, we would obtain the exact same shape as in Figure 1.C.1, which confirms the extreme variability of the rBergomi variance process for large maturities. The roughness of the variance process is again clearly visible in Figure 1.13.

1.7 Conclusion of the fit

From the empirical analyses of the implied volatility surface, we can conclude that rough volatility models with the rough Heston and the rBergomi clearly outperform standard SV models in terms of goodness of fit of the observed term structure of ATM volatility skews. Therefore, with the lowest RMSE (0.0846) and a good fit of the implied volatility surface, we conclude that the rough Heston model is the best of the considered models at reproducing the observed risk-neutral data. Secondly, since the rough Heston exhibits a rough variance process (with a H close to the historical estimate), this model can also better reproduce at any time scales numerous statistical properties of the historical volatility process. Therefore, the fact that this model exhibits rough volatility and hence better reproduces the empirical properties of the historical volatility will lead to a better pricing of insurance contracts at large time scales. More precisely, the parameter H (and $\lambda = 0$) controlling the roughness/memory of the variance process has a direct impact on long-term life insurance contracts since the generated sample paths of the fund's underlying price process S will strongly vary in function of this index H (as shown below in Section 1.8 and Appendix 1.B). In fact, the larger the maturity, the more important will be the effect of the roughness of the variance process on the fair value of such contracts. Moreover, as shown in Section 1.4.3, we can build at large time scales a stable regime for the variance process in the rough Heston, which ensures a reasonable long-term behavior of the model. Combined with the fact that the SSVI forward variance curve $\xi_0^{SSVI}(t)$ is consistent with the absence of arbitrage at any time scales, we will obtain in the next section a very reliable and consistent methodology for pricing long-term life insurance contracts. Finally, the rough Heston is also particularly tractable (only three parameters to calibrate, a characteristic function available in quasi-closed form and a time-efficient simulation scheme), which is very important in practice for fast calibration and option pricing.

Concerning the rBergomi model, even with calibrated parameters, it provides a slightly higher RMSE than the rough Heston model and hence a less good fit of the observed option prices (at least in our implementation with the SSVI forward variance curve). Secondly, the rBergomi does not provide realistic variance sample paths in the long run due to the huge level of the vol-of-var parameter and an exponentially increasing quantity $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$, leading to extreme values for large maturities. Furthermore, this model is fairly slow in terms of running time, which appears problematic for calibration and option pricing. To quote El Euch et al. (2019): *"Even with the introduction of the efficient hybrid scheme of Bennedsen et al. (2017), practical implementation [of the rBergomi model] has proved to be difficult"*. One solution to improve the stability and efficiency of this pricing model is to use asymptotics such as in Fukasawa (2017), in Forde and Zhang (2017) or Bayer et al. (2019, a). Finally, Bayer et al. (2019, b) propose a neural network approach to calibrate more accurately the rBergomi and to approximate the implied

volatility surface. We will however not deepen these methods in this chapter.

Finally, this SSVI extrapolation of the forward variance curve allows a more consistent risk-neutral valuation of the financial part of life insurance contracts, compared with standard calibration methods in finance. Usually, classical models such as the B&S or stochastic volatility models are calibrated based on the observed short-term option quotes on the market (or equivalently on short-term implied volatilities). From these calibrated parameters, we can then use standard option pricers (such as FFT pricers or analytical formulae) for obtaining the price of long-term financial derivatives. However, the generated implied volatility surface at large maturities from these models is known to be not realistic (see Gatheral (2011) or Figure 3.7.3 in Chapter 3). Moreover, this way of extrapolating option prices at large maturities does not prevent any static arbitrage. By using instead this SSVI parametrization/extrapolation of the observed (short-term) implied volatility surface and then computing the associated forward variance curve, we can extrapolate this curve in an arbitrage free way while ensuring good properties of the associated variance process, as seen above. Since this procedure reduces the number of parameters to estimate (only 3 in the rough Heston), it is also much more tractable and more accurate, compared with standard calibration methods (see Section 1.4–1.5).

1.8 Life insurance contract valuation

Although all models are well calibrated and hence vanilla options have approximately the same prices under all models, exotic option prices can differ dramatically. It is important to point out that vanilla options determine the marginal distribution (at maturity T of the option), not the process. Indeed, the underlying fine-grain properties of the process have an important impact on path-dependent option prices. We now highlight the impact of exotic price ranges between our four calibrated models by valuating our endowment life insurance contract described in Section 1.2. This way, we also emphasize the impact of rough volatility models on long-term insurance contracts and compare more thoroughly the rBergomi model (1.3.4) and the rough Heston (1.3.7) in terms of long-term properties. For further reference on equity-linked endowment valuation, we refer to Bacinello et al. (2010).

1.8.1 Valuation

By the first Fundamental Theorem of Asset Pricing, the fair value at time 0 of the cash-flow F_t in (1.2.1) is given by its discounted expected value under a risk-neutral measure $\mathbb{P}^*$. However, this cash flow also depends on the path of the Poisson process N , unknown at time 0. We thus take the following expectation conditionally to $\mathcal{G}_0$ to obtain the fair value of our equity-linked endowment contract with maturity T

$$FV_0 = \mathbb{E} \left[(1 - N_T) e^{-rT} \mathbb{E}^{\mathbb{P}^*} [F_T | \mathcal{F}_0] + \int_0^T e^{-rt} \mathbb{E}^{\mathbb{P}^*} [F_t | \mathcal{F}_0] dN_t \middle| \mathcal{G}_0 \right].$$

The processes N and S being independent, we obtain that

$$FV_0 = {}_T p_x e^{-rT} \mathbb{E}^{\mathbb{P}^*}[F_T] + \int_0^T e^{-rt} \mathbb{E}^{\mathbb{P}^*}[F_t] {}_t p_x \mu_x(t) dt.$$

We now want to price this insurance contract with the four models we calibrated above (Heston, Bates, rBergomi and rough Heston). We can then use a Monte-Carlo approach to simulate their price process S (and variance process V) and take the average of the paid benefit under each simulation $j = 1, \dots, m$ in order to derive the expectation under $\mathbb{P}^*$, *i.e.* :

$$FV_0 = \frac{1}{m} \sum_{j=1}^m \left(e^{-rt_n} F_{t_n, j} \mathbb{1} \left\{ \sum_{i=0}^n \Delta N_{t_i, j} = 0 \right\} + \sum_{i=0}^n e^{-rt_i} \Delta N_{t_i, j} F_{t_i, j} \mathbb{1} \left\{ \sum_{k=0}^{i-1} \Delta N_{t_k, j} = 0 \right\} \right),$$

where the underlying price process in F_t is defined by the risk-neutral pricing equation of each model. The discretization of their price process and variance process is given as in the previous Section 1.6. Moreover, the use of a Poisson process above makes it easy to add the mortality effect in the simulations to compute FV_0 . Indeed, for each simulation $j = 1, \dots, m$, we generate $n = T/\Delta t = 5/0.001 = 5000$ Poisson random variables $\Delta N_{t_i, j}$ with parameter $\lambda = \mu_x(t_i) \Delta t$ for $i = 0, \dots, n$. For the first i where $\Delta N_{t_i, j} = 1$, we compute a death benefit $F_{t_i, j}$ and if $\Delta N_{t_i, j} = 0$ for all i , we compute a survival benefit $F_{t_n, j}$.

We consider the following contracts :

- A female policyholder aged $x = 50$ or $x = 65$ at time $t = 0$.
- We use the Belgian life table FR for the Makeham force of mortality $\mu_x(t)$.
- The initial amount F_0 invested in the fund equals 10 000 €.
- The policyholder pays a unique premium P at time $t = 0$. Note that we could have imposed in the following that $FV_0 = P = F_0 = 10\,000$ € so as to make the contract *fair* and then determine the corresponding κ_g , κ_m or π in (1.2.1) that verifies this condition. However, since we aim to compare the contract's value FV_0 between models, we prefer to fix arbitrary values for κ_g , κ_m and π and then derive the corresponding value of the contract.
- We then assume the following parameters when valuating the contract :

Table 1.1: Parameters of the equity-linked endowments.

N° of contract	Gender	Initial age	κ_g	κ_m	π	Maturity T
1)	Female	50	0.01	0.2	0.8	5
2)	Female	50	0.01	0.2	0.8	10
3)	Female	50	0.01	0.2	0.4	10
4)	Female	50	0.01	0.4	0.8	10
5)	Female	65	0.01	0.2	0.8	15

We then obtain in Table 1.2 the fair values of the endowments with our different models⁵ using Monte-Carlo simulations with $m = 200\,000$ and working with the SSVI forward variance curve $\xi_0^{SSVI}(t)$ for the rBergomi and rough Heston models.

Table 1.2: Fair value of the endowment contracts with $\xi_0^{SSVI}(t)$.

N° of contract	Heston	Bates	rBergomi	rough Heston
1)	14 017.18 €	13 442.93 €	12 365.27 €	12 363.09 €
2)	19 353.14 €	17 649.19 €	14 893.48 €	14 130.58 €
3)	14 980.08 €	14 291.13 €	12 916.11 €	12 755.05 €
4)	20 853.01 €	19 079.05 €	15 412.74 €	15 070.89 €
5)	24 909.56 €	21 770.53 €	16 846.11 €	15 425.45 €

We can compare Table 1.2 with the following Table 1.3 computed using the constant forward variance curve $\xi_0^{cst}(t) = 0.0925$.

Table 1.3: Fair value of the endowment contracts with $\xi_0^{cst}(t)$.

N° of contract	Heston	Bates	rBergomi	rough Heston
1)	14 017.18 €	13 442.93 €	12 527.35 €	12 836.3 €
2)	19 353.14 €	17 649.19 €	15 492.96 €	15 430.33 €
3)	14 980.08 €	14 291.13 €	13 261.24 €	13 386.86 €
4)	20 853.01 €	19 079.05 €	16 294.61 €	16 762.31 €
5)	24 909.56 €	21 770.53 €	17 925.66 €	17 513.7 €

We first see that the Heston model leads to a higher price than the Bates model in both tables. Indeed, even if the long-term mean variance θ^* is higher in the Bates model, the negative average jumps combined with a lower initial variance V_0^* and a lower speed of mean reversion λ^* lead to a lower price in the Bates model. Secondly, we have that rough-type models have a lower price than standard SV models. This can be partly explained in Table 1.2 by the long-term variance (equal on average in rough models to $\xi_0^{SSVI}(t)$), which is lower than the long-term mean θ^* in the Heston ($= 0.097$) and Bates model ($= 0.271$), as depicted on Figure 1.5. However, if we analyze Table 1.3 with $\xi_0^{cst}(t) = 0.0925$, we still have lower prices for rough volatility models even if the average long-term variance is this time comparable. This remaining difference can thus be explained by the rough properties of the variance process, and in particular the very long reversion time scale in rough volatility models ($\lambda \rightarrow 0$). Moreover, we can see in Table 1.2 that the rBergomi model tends to have a higher fair value than the rough Heston due to the extreme values of $\mathbb{V}^{\mathbb{P}^*}[V_t|\mathcal{F}_0]$ for large t and the lower Hurst index H . This is especially true at long time scales and when the participation rate π is high. Finally, comparing

⁵Note that we consider here the standard version of the Heston model. Similar results are given in Appendix 1.B.3 with the forward-variance specification of this Heston model.

the different contracts, we clearly see that the larger the maturity, the higher is the difference between rough volatility models and standard SV models. This will be confirmed with Table 1.5 below. Moreover, when the participation rate π decreases, the fair value of the contract is logically lower since the policyholder earn a lower portion of the yearly returns. This lower π also leads to less pronounced differences between rough and standard SV models since their divergent sample paths properties play a lesser role. We observe the same effect when the maximal yearly return κ_m increases, with higher differences in fair value between rough and standard SV models due to their divergent sample paths.

Since we said that the rough Heston model is the most consistent with risk-neutral data while enjoying the best long-term properties, we retain the fourth column of Table 1.2 as being the contracts' prices which are the most market-consistent. We confirm these effects with the two following tables, displaying the fair value FV_0 in function of the minimum yearly guaranteed rate k_g and the maturity of the contract T . These two tables are computed using the SSVI forward variance curve but similar tables derived from $\xi_0^{cst}(t)$ and from $\xi_0^{mkt}(t)$ can be found in Appendix 1.B with Tables 1.B.1 and 1.B.2 and Tables 1.B.3 and 1.B.4, respectively. These tables in Appendix 1.B are also consistent with the following analyses.

Table 1.4: Fair value FV_0 for various k_g with maturity $T = 10$ and with $\xi_0^{SSVI}(t)$.

$T = 10$	Heston	Bates	rBergomi	rough Heston
$k_g = 0.5\%$	18 931.16 €	17 255.14 €	14 494.43 €	13 696.86 €
$k_g = 1\%$	19 353.14 €	17 649.19 €	14 893.48 €	14 130.58 €
$k_g = 2\%$	20 262.77 €	18 498.54 €	15 795.31 €	15 130.80 €
$k_g = 5\%$	23 644.91 €	21 649.22 €	19 418.47 €	19 072.87 €

We observe the highest differences between rough and standard SV models when k_g tends to be small. Indeed, in this case, the minimum yearly guarantee is exercised less often and therefore the yearly returns, which are highly influenced by the sample paths, play a more important role. This thus leads to stronger differences in fair value between models with divergent sample path properties.

Table 1.5: Fair value FV_0 for various maturities T with $k_g = 1\%$ and with $\xi_0^{SSVI}(t)$.

$k_g = 1\%$	Heston	Bates	rBergomi	rough Heston
$T = 5$	14 017.18 €	13 442.93 €	12 365.27 €	12 363.09 €
$T = 10$	19 353.14 €	17 649.19 €	14 893.48 €	14 130.58 €
$T = 20$	35 618.91 €	29 613.34 €	20 126.21 €	17 677.04 €

The differences in fair value between models are also the most exacerbated when the maturity of the contract is large. The impact of the sample paths properties of

each model is indeed stronger at longer time scales. In particular, the differences in the long-term behavior of the variance process are exacerbated when the maturity is large, as explained in the previous sections and in Appendix 1.C.2. Combined with the fact that rough volatility models better reproduce the statistical properties of the historical volatility thanks to their memory properties, this confirms the importance and the added value of using rough volatility models at large time scales, and especially the rough Heston model. Particularly, the importance of the Hurst index (controlling the roughness and memory of the variance process) is emphasized in Appendix 1.B when comparing the fair values obtained with the different forward variance curves and with the forward-variance specification of the Heston model. We indeed show that this Hurst index has a huge impact on the contracts' fair values, especially for large maturities. We finally want to highlight that the numbers given in the tables above can slightly change for each run of simulations but are quite stable with our chosen m , equal to 200 000.

1.9 Conclusion

Rough fractional stochastic volatility models are excellent candidates for reproducing important stylized facts of the past volatility time series and for providing a consistent implied volatility surface with the observed one. In this first empirical chapter, we explore more closely two tractable implementations of the RFSV model, the rBergomi and the rough Heston models. We show how to expand these two rough volatility models at larger time scales, we analyze their implications for the pricing of long-term life insurance claims and we explain why they provide a more accurate fair value of such long-term contracts. In particular, we study the long-term properties of these two rough volatility models and compare them with two standard SV models (the Heston and Bates models).

Among these four models, we conclude that the rough Heston is the most consistent with historical and risk-neutral data while enjoying the best long-term properties. Indeed, this rough Heston model with $H \approx 0.15$ tends to outperform standard SV models in terms of calibration with a better fit of the implied volatility surface and allows to better reproduce the statistical properties of the observed historical volatility, especially at long time scales. We also retain the rough Heston model as a highly tractable implementation of rough volatility models with only three parameters to estimate and a characteristic function available in quasi-closed form, which is not the case of the rBergomi model. Combined with the Carr-Madan formula, the calibration and pricing of European options with the rough Heston model is hence extremely fast and accurate. Moreover, we manage to build at large time scales a stable regime for the variance process in the rough Heston based on a SSVI parametrization of the forward variance curve. This way, we obtain a model with a reasonable long-term behavior for its variance process while being consistent with the absence of static arbitrage. This rough Heston is thus able to overcome

the extreme and exponentially increasing variability of the rBergomi variance process. Using the pricing methodology introduced in this chapter, a more accurate fair value of long-term equity-linked life insurance contracts is then obtained in comparison with the rBergomi and standard SV models. This is especially true at large maturity where the good properties of the rough Heston model have a stronger impact. More precisely, we obtain with the rough Heston model a fair value significantly lower than in standard SV models due to the long-reversion time scale of volatility. This highlights the added-value of using this rough Heston model when pricing long-term life insurance contracts.

Note that we do not claim in this chapter that the historical volatility is rough but we simply show that rough volatility models (and in particular the rough Heston) with Hurst index below $1/2$ best fit the observed option quotes in the market and the associated volatility surface (see Section 1.4 and 1.5). In fact, to quote Cont (2007): *"The econometric debate on the short range or long range nature of dependence in volatility still goes on (and may probably never be resolved)"*. And Gatheral et al. (2018) pursue: *"In our view, the question as to whether the volatility time series exhibits long memory (in a technical sense) or not is not a very useful or fruitful one."*

As future work, we should re-calibrate all the models considered in this first chapter on other market data of European options. Choosing an other index/stock with a larger set of available maturities and at another point in time should help us to confirm and validate the results derived with our methodology. Moreover, more advanced mortality models could be considered such as in Hainaut and Devolder (2008) or in Cairns et al. (2008). Finally, the methodology introduced in this paper based on a SSVI interpolation and extrapolation of the observed forward variance curve could be applied to any pricing model in forward variance form. A more thorough comparison study of life insurance pricing should hence be carried out based on this approach.

1.A SSVI forward variance curve

The surface SVI (SSVI) introduced in Gatheral and Jacquier (2014) provides a tractable arbitrage-free parametrization of the implied volatility surface. SSVI is based on a simple closed-form representation of this volatility surface, which allows an extremely efficient calibration of this parametrization to observed implied volatility data. For a more detailed survey on such surface SVI, we refer to Gatheral and Jacquier (2014). We simply recall here that SSVI parametrizes the total implied variance (usually denoted $w(k, \theta_t)$) in terms of the log-strike k , of the ATM total variance $\theta_t := \sigma_{imp}^2(0, t)t$ (which is assumed to be read on the market), of a constant leverage parameter ρ and of a curvature function $\varphi(\theta_t)$. We will consider the widely-used power-law curve for $\varphi(\cdot)$,

$$\varphi(\theta_t) = \frac{\eta}{\theta_t^\gamma (1 + \theta_t)^{1-\gamma}},$$

since it allows a SSVI volatility surface completely free of static arbitrage (no calendar spread and butterfly arbitrages) provided that $\gamma \in (0, 1/2]$ and $\eta(1+|\rho|) \leq 2$. This function also allows to be consistent with the power-law shape of the observed term structure of ATM volatility skew (see Section 1.5).

We then use the algorithm described in Section 5.2 of Gatheral and Jacquier (2014) with our power-law SSVI in order to calibrate the parameters (ρ, η, γ) to implied volatility data while avoiding static arbitrage. The optimal parameters ρ^*, η^* and γ^* are equal respectively to $(-0.6842, 1.1536, 0.3410)$ in such a way that the non-arbitrage conditions above are verified. With these parameters, we obtain Figure 1.A.1 below depicting the excellent fit of the power-law SSVI parametrization to observed implied volatility data. We observe that the quality of the fit is almost perfect. There is only a slight deviation from the data for the shortest maturity $T = 0.027$. The associated forward variance curve to this SSVI parametrization is then given by Figure 1.A.2 and is again obtained as in Gatheral (2011) by approximating the value of a variance swap with the value of a log-contract of the same maturity, built this time upon the SSVI parametrization of the volatility surface. Note that this SSVI forward variance curve is interpolated between observed maturities such as to ensure the absence of arbitrage in the interpolated volatility surface, following the methodology described in Section 5.3 of Gatheral and Jacquier (2014).

The objective now is to extrapolate in an arbitrage-free way the SVI surface beyond the last available maturity $T = 1.2$. As explained again in Section 5.3 of Gatheral and Jacquier (2014), we first need a monotonic increasing extrapolation of θ_t which is asymptotically linear in time. Based on Figure 1.A.3 below depicting the values of θ_t for the observed maturities, we clearly see that a simple linear regression appears to provide the best fit (while being of course asymptotically linear

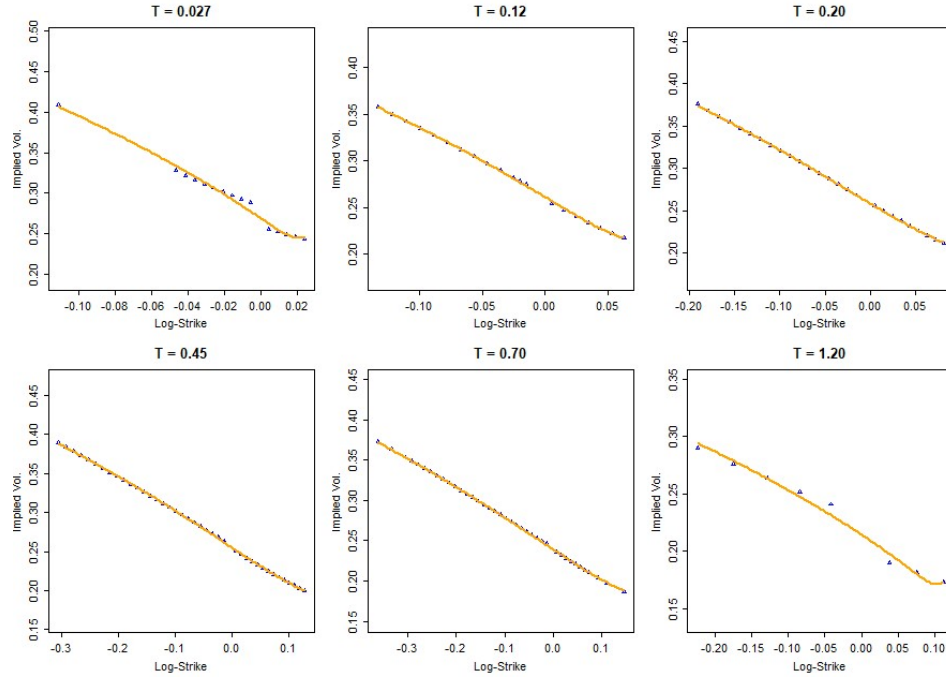


Figure 1.A.1: Blue dots are the observed mid implied volatilities, the orange solid line is the power-law SSVI fit.

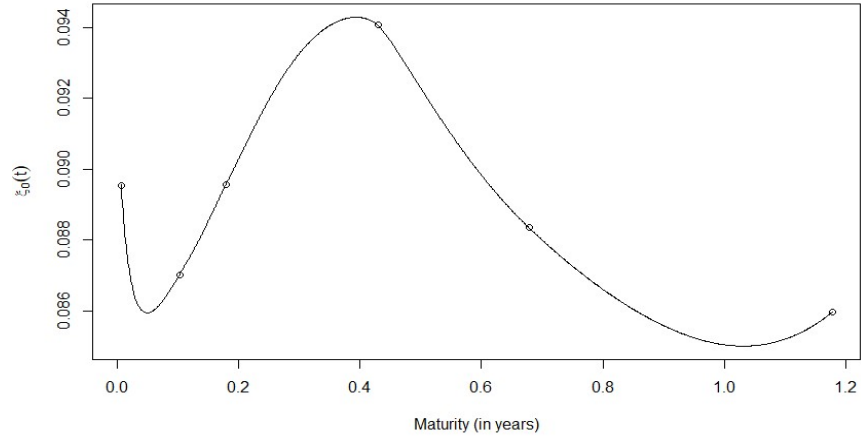


Figure 1.A.2: Initial SSVI forward variance curve $\xi_0^{SSVI}(t)$ between 0 and 1.2 years. The black circles are the values of $\xi_0^{SSVI}(t)$ for the observed maturities.

in time). Let t_n be the last available maturity in the data. According to Gatheral and Jacquier (2014), the total implied variance for $t > t_n$ can now be extrapolated by

$$w(k, \theta_t) = w(k, \theta_{t_n}) + \theta_t - \theta_{t_n},$$

while remaining free of static arbitrage (see Theorem 4.3 in Gatheral and Jacquier (2014)). Finally, the forward variance curve is again retrieved for $t > t_n$ from the extrapolated values $w(k, \theta_t)$ by valuating log-contracts. The extrapolated forward variance curve based on this power-law SSVI parametrization is depicted in Figure 1.5, see Section 1.4.3.

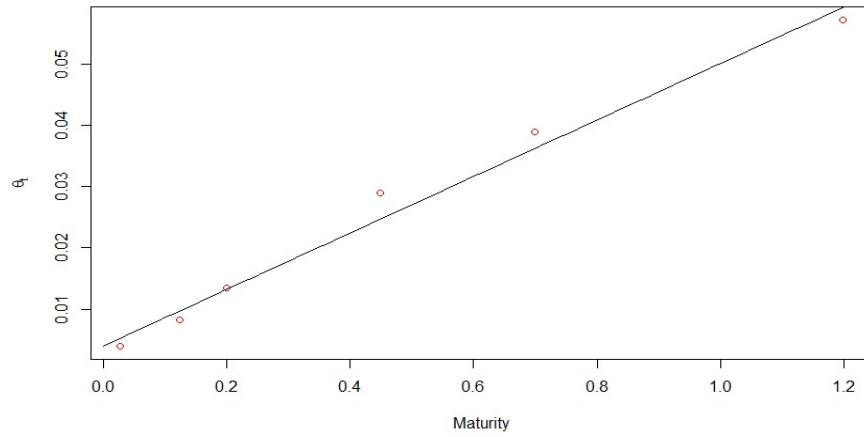


Figure 1.A.3: Red dots are values of θ_t for the available maturities, the red line is the linear fit.

1.B Calibration and valuation based on other forward variance curves

1.B.1 Constant forward variance curve

Rough Heston :

Using the constant forward variance curve $\xi_0^{cst}(t) = 0.0925$, we obtain in a first step the following set of initial parameters from the SA global optimizer : (0.1235, 0.6383, -0.6415). Using our local optimizer, we then find a minimal RMSE of 0.11097 with the following (same) final parameter values Θ^* :

H^*	ν^*	ρ^*
0.1235	0.6383	-0.6415

The fit is thus poorer than with the initial SSVI forward variance curve $\xi_0^{SSVI}(t)$. The vol-of-var parameter ν^* and the anti-correlation ρ^* are close to values obtained previously with $\xi_0^{SSVI}(t)$ but the Hurst index H^* is now lower and closer to the historical estimate ($= 0.13$).

rBergomi :

Using $\xi_0^{cst}(t) = 0.0925$, we find a RMSE of 0.0851 with the initial set of parameters (0.12, 1.55, -0.90). This leads to the following final parameter values Θ^* :

H^*	η^*	ρ^*
0.1231	1.8636	-0.8951

Hence, we obtain a slightly better fit than with the rBergomi based on the SSVI forward variance curve (and almost as good as the SSVI rough Heston). However, we have an even higher vol-of-var parameter compared with the SSVI forward variance curve

$$\nu^* = \eta^* \sqrt{2H^*} / (2C_H) = 1.137$$

This huge level ν^* again leads to unrealistic sample paths. We also obtain a lower Hurst index H^* . Moreover, we still have a stability issue when calibrating this rBergomi model with a high variability of the results in function of the chosen number of paths and steps (here, $m = 1\,000\,000$ and $n = 300$ for the calibration).

Furthermore, the two Tables 1.B.1 and 1.B.2 display the fair value FV_0 in function of the minimum yearly guaranteed rate k_g and the maturity of the contract T .

Table 1.B.1: FV_0 for various k_g with maturity $T = 10$ and $\xi_0^{cst}(t)$

$T = 10$	rBergomi	rough Heston
$k_g = 0.5\%$	15 086.55 €	15 041.99 €
$k_g = 1\%$	15 492.96 €	15 430.33 €
$k_g = 2\%$	16 403.19 €	16 334.91 €
$k_g = 5\%$	20 015.25 €	20 048.91 €

Table 1.B.2: FV_0 for various maturities T with $k_g = 1\%$ and $\xi_0^{cst}(t)$

$k_g = 1\%$	rBergomi	rough Heston
$T = 5$	12 527.35 €	12 836.30 €
$T = 10$	15 492.96 €	15 430.33 €
$T = 20$	22 079.51 €	21 124.48 €

If we compare these two tables with Tables 1.4 and 1.5 derived from the SSVI forward variance curve, we see that the obtained fair values are significantly higher

when using the constant forward variance curve $\xi_0^{cst}(t)$. For the rough Heston, we said above that the main difference in terms of calibration is the lower Hurst index and the higher long-term mean of V when considering $\xi_0^{cst}(t)$ instead of $\xi_0^{SSVI}(t)$, leading to a higher fair value of the contracts. Indeed, the lower the Hurst index, the higher the fair value. Moreover, we observe that the differences in fair values when considering the two forward variance curves are the most exacerbated at large time scales. This comparison hence emphasizes the importance of taking the roughness/memory property of the variance process into account, especially for large maturities. Finally, we again have in Table 1.B.2 with $\xi_0^{cst}(t)$ that the rBergomi model leads to higher fair values than the rough Heston at large maturities due to the explosion of $\mathbb{V}^{\mathbb{P}^*}[V_t|\mathcal{F}_0]$ (despite the higher ν^* and the lower ρ^* in the calibration).

1.B.2 Market forward variance curve

Rough Heston :

Using the market forward variance curve $\xi_0^{mkt}(t)$, we obtain in a first step the following set of initial parameters from the SA global optimizer : (0.1206, 0.4131, -0.8015). We then obtain with our local optimizer a minimal RMSE of 0.09235 with the following (same) final parameter values Θ^* :

H^*	ν^*	ρ^*
0.1206	0.4131	-0.8015

The fit is thus poorer compared with the fit derived from the SSVI forward variance curve but better than when using $\xi_0^{cst}(t)$. The Hurst index is again lower and very close to the historical estimate. We also observe that the volatility of variance parameter ν^* is lower compared to the SSVI parametrization and we obtain this time a stronger anti-correlation ρ^* .

rBergomi :

We find a RMSE of 0.1479 with the initial set of parameters equal to (0.1, 1.45, -0.99). This leads to the following final values Θ^* :

H^*	η^*	ρ^*
0.1357	1.4614	-0.9967

The quality of the fit is clearly poorer compared with $\xi_0^{SSVI}(t)$ and $\xi_0^{cst}(t)$. The vol of var parameter is this time lower and more reasonable. We indeed have $\nu^* = \eta^* \sqrt{2H^*}/(2C_H) = 0.8804$. We again obtain a lower Hurst index compared with the SSVI forward variance curve.

Table 1.B.3: Fair value FV_0 for various k_g with maturity $T = 10$ and $\xi_0^{mkt}(t)$

$T = 10$	rBergomi	rough Heston
$k_g = 0.5\%$	14 518.48 €	13 817.37 €
$k_g = 1\%$	14 909.54 €	14 228.54 €
$k_g = 2\%$	15 785.42 €	15 196.15 €
$k_g = 5\%$	19 301.23 €	19 075.75 €

Table 1.B.4: Fair value FV_0 for various maturities T with $k_g = 1\%$ and $\xi_0^{mkt}(t)$

$k_g = 1\% \text{ s}$	rBergomi	rough Heston
$T = 5$	12 291.17 €	12 371, 10 €
$T = 10$	14 909.54 €	14 228.54 €
$T = 20$	20 639.81 €	18 293.85 €

Since the market forward variance curve lies below the SSVI forward variance curve for the considered maturities (see Figures 1.4-1.5), we would expect a lower fair value of our contract using $\xi_0^{mkt}(t)$ instead of $\xi_0^{SSVI}(t)$. However, for both the rough Heston and the rBergomi, the lower Hurst index H leads to an increase of the contract's prices and hence, it compensates for the lower average values of V . This is why we observe very similar fair values when using the two forward variance curves (cf. Tables 1.B.3 and 1.B.4 vs. Tables 1.4 and 1.5). For large maturities or small values of κ_g , we see that we even obtain a higher fair value when using the market forward variance curve $\xi_0^{mkt}(t)$. Again, this comes from the fact that the sample paths properties are exacerbated when the maturity is large or when the minimal guarantee κ_g is low. This comparison also shows that the Hurst index H (and hence the roughness/memory property of the variance process) has a strong impact on the fair value of insurance contracts, especially at large time scales. This emphasizes once again the importance of using rough volatility models.

1.B.3 Forward Heston model

In order to provide a better comparison with rough volatility models, we finally introduce the forward-variance specification of the Heston model. From Gatheral and Keller-Ressel (2019), we have that the Heston model is an affine forward variance model, as explained in Section 1.3.2. Hence, the characteristic function of the log-asset price $X_t = \log S_t$ seen from time 0 under the Heston model can be written as previously by (1.3.8) where we use the SSVI forward variance curve $\xi_0^{SSVI}(t)$ and where we need to solve the Riccati ODE (1.3.9). Using well-known numerical methods to solve this Riccati ODE and applying the Carr-Madan formula with the characteristic function above, it is quite simple to calibrate and find the optimal parameters $(\lambda^*, \rho^*, \nu^*)$ based on the observed European options and based on the SSVI forward variance curve $\xi_0^{SSVI}(t)$. We find a RMSE of 0.1182 with the following

optimal parameters :

λ^*	ν^*	ρ^*
3.02522	1.42928	-0.71762

The RMSE and the final parameters of this forward-variance specification of the Heston model are highly similar to the ones derived above in Section 1.4.2 for the classical specification of the Heston model. However, for pricing our life-insurance contract, we still need to calibrate the mean-reversion level θ and the initial variance level V_0 . The constant parameter θ is not flexible enough to deal with a given forward variance curve and we therefore need to consider an extension of the Heston model where we allow the mean-reversion level $\theta(t)$ to be time-dependent. In order to fit the market and recover the SSVI forward variance curve, Buehler (2006) show that we must set

$$\theta(t) = \frac{\frac{\partial}{\partial t} \xi_0^{SSVI}(t)}{\lambda} + \xi_0^{SSVI}(t) \quad \text{and} \quad V_0 = \xi_0^{SSVI}(0).$$

This indeed allows to also have $\mathbb{E}^{\mathbb{P}^*}[V_t|\mathcal{F}_0] = \xi_0^{SSVI}(t)$ in this model. We then obtain the following Tables 1.B.5 and 1.B.6 giving the fair value FV_0 of the contracts above under the forward-variance specification of the Heston model.

Table 1.B.5: FV_0 for various k_g with maturity $T = 10$ and $\xi_0^{SSVI}(t)$.

$T = 10$	Heston	fwd Heston	rBergomi	rough Heston
$k_g = 0.5\%$	18 931.16 €	17 370.17 €	14 494.43 €	13 696.86 €
$k_g = 1\%$	19 353.14 €	17 750.09 €	14 893.48 €	14 130.58 €
$k_g = 2\%$	20 262.77 €	18 581.37 €	15 795.31 €	15 130.80 €
$k_g = 5\%$	23 644.91 €	21 790.04 €	19 418.47 €	19 072.87 €

Table 1.B.6: FV_0 for various maturities T with $k_g = 1\%$ and $\xi_0^{SSVI}(t)$.

$k_g = 1\%$	Heston	fwd Heston	rBergomi	rough Heston
$T = 5$	14 017.18 €	13 442.93 €	12 365.27 €	12 363.09 €
$T = 10$	19 353.14 €	17 750.09 €	14 893.48 €	14 130.58 €
$T = 20$	35 618.91 €	28 928.57 €	20 126.21 €	17 677.04 €

We clearly see that the forward-variance specification of the Heston model provides lower fair values compared with the standard Heston model due to the lower mean-reversion level. The mean-reversion function $\theta(t)$ indeed lies below the value $\theta^* = 0.0967$ of the standard Heston model at any time scales. We again see that this forward-variance Heston model provides a higher fair value than rough volatility models even if their average long-term variance is this time comparable since

$\mathbb{E}^{\mathbb{P}^*}[V_t|\mathcal{F}_0] = \xi_0^{SVI}(t)$ in each model. This once again emphasizes that the rough property of the variance process induced by the index H^* and $\lambda \rightarrow 0$ has a strong impact on the obtained fair values of long-term life insurance contracts.

1.C Figures

1.C.1 Risk-free rates

Table 1.C.1: European risk-free interest rates between 0 and 20 years.

Maturity	Rate	Maturity	Rate
Overnight	-0.55486 %	8 years	-0.26535 %
7 days	-0.52686 %	9 years	-0.23026 %
1 month	-0.494 %	10 years	-0.19118 %
2 months	-0.44157 %	11 years	-0.15312 %
3 months	-0.43514 %	12 years	-0.11707 %
6 months	-0.35386 %	13 years	-0.08303 %
1 year	-0.36767 %	14 years	-0.07203 %
2 years	-0.38775 %	15 years	-0.02600 %
3 years	-0.39076 %	16 years	-0.00200 %
4 years	-0.37771 %	17 years	0.00099 %
5 years	-0.35563 %	18 years	0.0020 %
6 years	-0.33054 %	19 years	0.01100 %
7 years	-0.29945 %	20 years	0.03599 %

1.C.2 Conditional variance process

The rBergomi variance process is conditionally log-normal and we can prove that

$$\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t] = \xi_t(u)^2 \left(\exp(\eta^2(u-t)^{2H}) - 1 \right). \quad (1.C.1)$$

Hence, this quantity increases exponentially with u , leading to extreme swings and variability in the rBergomi variance process for large maturities. We see that $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$ also depends on the forward variance curve, but since this curve is stable for large maturity (see Figure 1.5 and 1.4), it does not contribute to the exponential growth of $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t]$.

For the rough Heston, using Itô isometry and conditional log-normality, we find

$$\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t] = \frac{\nu^2}{[\Gamma(H+1/2)]^2} \int_t^u (u-s)^{2H-1} \xi_t(s) ds. \quad (1.C.2)$$

Since the forward variance curve is almost constant for large maturities, we make the assumption that this curve is fixed to a constant K so as to obtain the following

approximation at $t = 0$:

$$\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_0] \approx \frac{K \nu^2}{[\Gamma(H + 1/2)]^2 2H} u^{2H}.$$

This approximation allows to see that the rate of increase of the quantity $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_0]$ is approximately u^{2H} , which leads to a slowly increasing behavior as confirmed on Figure 1.C.1. Using the SSVI forward variance curve $\xi_0^{SSVI}(u)$ in equation (1.C.1) with $\eta = 1.6803$ and $H = 0.1515$ and using again $\xi_0^{SSVI}(u)$ with $\nu = 0.6241$ and $H = 0.1703$ in equation (1.C.2), we obtain the following figure Figure 1.C.1 at $t = 0$.

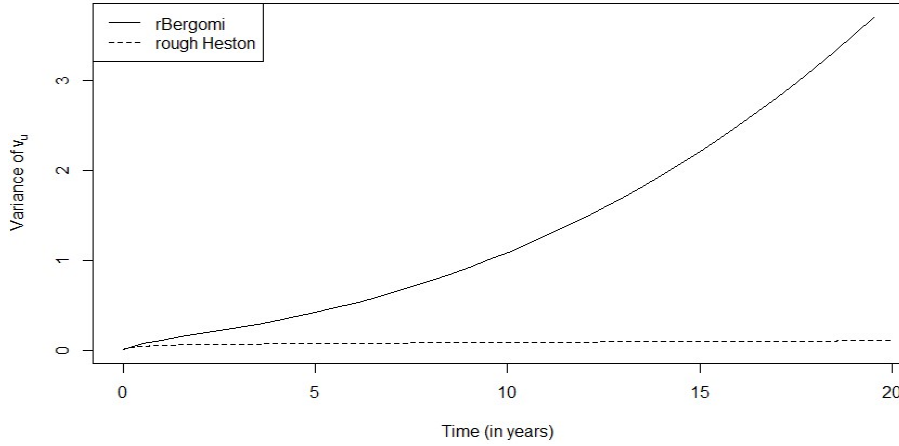


Figure 1.C.1: Comparison of $\mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_0]$ in function of u between the rough Heston and the rBergomi models.

Finally, note that we indeed have $\lim_{u \rightarrow \infty} \mathbb{V}^{\mathbb{P}^*}[V_u|\mathcal{F}_t] = +\infty$ in both models, which confirms the non-stationary property of the rBergomi and rough Heston's variance processes, see Chen et al. (2011).

CHAPTER 2

Portfolio insurance under rough volatility and Volterra processes

Affine Volterra processes have gained increasing interest in recent years. In particular, this class of processes generalizes the classical Heston model and the more recent rough Heston model introduced in Chapter 1. The aim of this chapter is hence to revisit and generalize the constant proportion portfolio insurance (CPPI) under affine Volterra processes. Indeed, existing simulation-based methods for CPPI do not apply easily to this class of processes. We instead propose an approach based on the characteristic function of the log-cushion which appears to be more consistent, stable and particularly efficient in the case of affine Volterra processes compared with the existing simulation techniques. Using such approach, we describe in this chapter several properties of CPPI which naturally result from the form of the log-cushion's characteristic function under affine Volterra processes. This allows to consider more realistic dynamics for the underlying risky asset in the context of CPPI and hence build portfolio strategies that are more consistent with financial data. In particular, we address the case of the rough Heston model, known to be extremely consistent with past data of volatility as shown in Chapter 1, and provide a new estimation procedure for its parameters based on the PMCMC algorithm. Finally, the last part of the chapter investigates gap risk in discrete-time CPPI, which arises from liquidity constraints and transaction costs. Therefore, we manage in this chapter to study more accurately the true properties of such CPPI strategy and to better handle the (illiquidity) risk associated with it.

Dupret, J. L., & Hainaut, D. (2021). Portfolio insurance under rough volatility and Volterra processes. *International Journal of Theoretical and Applied Finance*, 24(06n07), 2150036.

2.1 Introduction

The constant proportion portfolio insurance (CPPI) is a dynamic strategy of investment that yields a super-linear participation in future asset returns while retaining a security guarantee on a part of the invested capital (called the floor). Under the assumption of frictionless trading, this strategy therefore protects a portfolio of stocks against the downside risk but still allows for an upside potential. Basically, the CPPI method consists in maintaining the risk exposure equal to a constant multiple of the excess of wealth over a floor. The CPPI was introduced for a portfolio of bonds by Perold and Sharpe (1988) and revisited by Black and Jones (1987) for equities. A first stream of literature extensively studies the theoretical properties of continuous-time CPPI strategies. A comparison of OBPI (option-based portfolio insurance) and CPPI is provided in Bertrand and Prigent (2001) under the classical Black and Scholes (B&S) framework and under the assumption of frictionless trading. This first set of research also deals with the effects of jump processes, stochastic volatility models, regime switching models and gap risk on the CPPI method, see Cont and Tankov (2009), Prigent and Bertrand (2003), Hainaut (2011), Balder et al. (2009) and references therein. The second stream of literature focuses on solving an optimization problem that arises when we assume investors maximize their end-of-period expected utility. In this setting, the optimality of the CPPI strategy depends on the risk profile of the investor, which has been thoroughly studied in Kingston (1989), Grossman and Vila (1992), Branger et al. (2010), Bertrand and Prigent (2019). Finally, the third stream tries to identify problems with the design of CPPI strategies and then proposes modifications to mitigate them. We refer to Pain and Rand (2008) for a summary of the different variations of CPPI proposed in the literature.

Our contribution lies within the framework of the first stream of literature. In this chapter, we derive a methodology based on the characteristic function of the log-cushion, which enables to derive and study the properties of continuous-time CPPI with frictionless trading under the very general class of affine Volterra processes, as defined in Abi Jaber et al. (2019). In fact, we show that the characteristic function of the log-cushion for affine Volterra processes is obtained in quasi-closed form, which then allows to easily and efficiently compute the moments, the density as well as several risk-measures for the CPPI value at maturity. Our approach is therefore not restricted to the study of CPPI properties under a particular model as already done in the literature but extends directly and easily to a wide class of models, allowing to take into account different market behaviors and more realistic dynamics for the underlying risky asset in the context of CPPI. Moreover, using the log-cushion's characteristic function instead of the existing simulation-based techniques allows to derive far more rapidly and more accurately the desired portfolio properties.

In particular, we will focus our analysis on CPPI properties under the rough Heston model, which is an important member among the class of affine Volterra processes (based on a power-law kernel), enjoying growing popularity among practitioners and academicians. Indeed, the rough Heston model of El Euch and Rosenbaum (2019) is a highly tractable implementation of rough volatility models which has been shown to be remarkably consistent with financial time series data, as initially explained in Gatheral et al. (2018). The rough Heston model therefore allows to better assess the true properties of the CPPI strategy compared with other dynamics for the underlying asset. More precisely, we will demonstrate that thanks to its rough variance process, the rough Heston model will lead to a more accurate assessment of the risk of the CPPI strategy by providing a better modeling of the left tail of the portfolio distribution. Indeed, as explained in Paulot and Lacroze (2009), this left tail must be computed with great precision since it plays an important role in the riskiness of such CPPI strategy. However, the non-Markovian nature of the variance process under the rough Heston model requires to deviate from the classical stochastic volatility approach for CPPI described in Prigent and Bertrand (2003) and to work with the characteristic function of the log-cushion instead. Moreover, the analysis of CPPI properties (performance, density and risk measures) will be carried out for the classical Heston model as well, which also belongs to the class of affine Volterra processes and for which well-established simulation techniques exist. Finally, we will compare the obtained results with the B&S model, which has already been extensively studied in the context of CPPI. In the last practical part of the chapter, we will show how to introduce gap risk and transaction costs in such CPPI strategy based on affine Volterra processes and hence go beyond the assumption of frictionless/continuous trading.

At last, we contribute with this work to provide a full econometric estimation of the rough Heston's parameters based on the time series of observed log-returns using the Particle Monte-Carlo Markov Chain algorithm. So far, this rough Heston model has only been calibrated in the literature based on the risk-neutral option quotes observed on the market and we now propose a new estimation procedure for deriving these parameters directly and efficiently under the real-world probability measure, which is required in the case of portfolio insurance strategies. This way, we will also show that the rough Heston model provides a better statistical fit of past log-return data compared with the Heston and B&S models.

2.2 Model setup under rough volatility

As mentioned above, the CPPI strategy consists in maintaining the risk exposure equal to a constant multiple m of the excess of wealth over a floor F . This way, the value of the portfolio $P = (P_t)_{t \geq 0}$ is above the floor F at any time t in the considered period of time $[0, T]$. The value of the floor thus gives the dynamical

insured amount and is assumed to evolve at a constant interest rate r following

$$dF_t = F_t r dt.$$

Note that a more complex dynamic could be chosen for modeling the interest rates but this choice does not impact the conclusions of our work. Obviously, the initial floor F_0 is less than the initial portfolio value P_0 and is equal to $F_0 = e^{-rT}G$, where G is the guaranteed amount at maturity. The difference $P_0 - F_0$ is called the cushion, denoted by C_0 . Its value C_t at any time $t \in [0, T]$ is given by

$$C_t = P_t - F_t.$$

Now let us denote the exposure E_t , which is the total amount invested in the risky asset S_t . The standard CPPI method consists of letting $E_t = mC_t$ where m is the constant multiple. In order to have a convex payoff in the underlying, we impose $m > 1$. From Bertrand and Prigent (2016), we have that m is an exogenous parameter chosen by the portfolio manager. A first alternative is to choose m as the inverse of the maximal negative daily return anticipated by the fund manager. For example, if the manager expects that the worst daily return will be -25% over $[0, T]$, then the upper bound on the multiple m such that the value of the fund P_t remains above the floor F_t for all time $t \in [0, T]$ is given by 4. A more accurate alternative is to choose m according to the occurrence probabilities of extreme events in the risky asset, such as described in Bertrand and Prigent (2016). We will apply their methodology in Section 2.6 to provide an upper bound of the multiple m based on a dataset of S&P500 daily log-returns.

Moreover, we consider in this chapter a continuous-time CPPI strategy and we assume in a first step continuous/frictionless trading over the whole investment horizon $[0, T]$. In the last section 2.8, we will then briefly discuss the possible ways of taking into account gap risk, illiquidity constraints and transaction costs by restricting trading on a discrete time grid.

We also assume in the sequel that the risky asset $S = (S_t)_{t \geq 0}$ is a diffusion process such that

$$dS_t = S_t \left(\mu dt + \sqrt{V_t} dW_t \right), \quad (2.2.1)$$

where W is a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a right-continuous filtration $(\mathcal{F}_t)_{t \geq 0}$. The cornerstone of this section is to consider that the variance process $V = (V_t)_{t \geq 0}$ exhibits a rough behavior. Indeed, in their groundbreaking paper, Gatheral et al. (2018) show that for a very wide range of assets, historical volatility time series exhibit a behavior that is much rougher than that of a Brownian motion (as implied by standard stochastic volatility models). More precisely, they emphasize that volatility dynamics based on fractional Brownian motions with a Hurst exponent in $(0, \frac{1}{2})$ are extremely consistent with the historical volatility process of major indices. Such condition on the Hurst index

aims to generate rough sample paths and to introduce short-range dependence in the volatility process, as explained in Chapter 1. On the contrary to long-memory volatility models ($H > 1/2$), this rough behavior of volatility allows to perfectly reproduce at any time scales a lot of statistical properties of the past volatility time series, as confirmed in many other studies such as Bennedsen et al. (2016). Furthermore, Gatheral et al. (2018) find that classical statistical procedures aiming at detecting volatility persistence tend to conclude the presence of long memory in data generated from their rough volatility model. This allows these authors to explain and understand why long memory of volatility has been widely accepted as a stylized fact until now. This recent change of paradigm towards rough volatility, both by academicians and practitioners, allows to study more accurately the true properties of CPPI strategies. Indeed, such comparisons with other portfolio insurance techniques have been mainly limited so far to the B&S case or to classical stochastic volatility models based on Brownian motions. We will show in this chapter the impact of taking into account rough volatility on CPPI properties by considering the rough Heston model of El Euch and Rosenbaum (2018) and we will explain why it provides a better modeling of the risk associated with the CPPI strategy. This rough Heston model is in fact an extension of the classical Heston model where El Euch and Rosenbaum (2018) introduce the power-law kernel $(t-u)^{H-1/2}$ in the Heston variance process V in order to allow for a rough behavior of V in a Heston-type model. They hence find the following variance process¹ for the rough Heston model under $\mathbb{P}$:

$$V_t = V_0 + \frac{1}{\Gamma(H + 1/2)} \int_0^t \frac{\lambda(\theta - V_u)}{(t-u)^{1/2-H}} du + \frac{\nu}{\Gamma(H + 1/2)} \int_0^t \frac{\sqrt{V_u}}{(t-u)^{1/2-H}} d\widehat{W}_u, \quad (2.2.2)$$

where λ is the speed of mean reversion of the variance process towards the level θ and where ν is the volatility of variance parameter. $\widehat{W}$ is again a standard Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\langle W, \widehat{W} \rangle_t = \rho t$ with a constant correlation ρ . When $H = 1/2$, we can verify that we indeed recover the classical Heston model. It can be shown that the trajectories of the variance itself under the rough Heston are almost surely Hölder-continuous of order $H - \varepsilon$, for any $\varepsilon > 0$, leading to a rough behavior of the variance process with short-range memory when $H < 1/2$ and smoother trajectories with long-range dependence when $H > 1/2$. These authors also provide in El Euch and Rosenbaum (2019) a microstructure foundation of such model as arising from the high-frequency modeling of market liquidity.

We define in this chapter the forward variance curve $\xi_s(t) := \mathbb{E}^\mathbb{P}[V_t | \mathcal{F}_s]$, $t \geq s$. We can easily recover V_s from $\xi_s(t)$ as $V_s = \lim_{t \rightarrow s} \xi_s(t)$. Following Bergomi and Guyon (2012), forward variance models are models that can be written as a function of this curve $\xi_s(t)$. As shown in El Euch and Rosenbaum (2018), this is the case of the rough Heston model (2.2.2). More precisely, we have the following

¹Note that compared to equation (1.3.6) in Chapter 1, the rough Heston model is given here under the physical measure $\mathbb{P}$ and with a constant mean-reversion level θ .

forward variance curve dynamic for the rough Heston model under $\mathbb{P}$:

$$d\xi_s(t) = \nu(t-s)^{\alpha-1} E_{\alpha,\alpha}(-\lambda(t-s)^\alpha) \sqrt{V_s} d\widehat{W}_s, \quad (2.2.3)$$

where $\alpha = H + 1/2$ and $E_{\alpha,\beta}(z)$ the generalized Mittag-Leffler function defined by

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

Let now $\eta_s(t) = \sqrt{V_s} \kappa(t-s)$ where $\kappa(t-s) = \nu(t-s)^{\alpha-1} E_{\alpha,\alpha}(-\lambda(t-s)^\alpha)$, we can then rewrite (2.2.3) as

$$d\xi_s(t) = \eta_s(t) d\widehat{W}_s. \quad (2.2.4)$$

This writing will be useful in the sequel to ease notations. Note that $\eta_s(t)$ satisfies the Assumption 2.1. of Gatheral and Keller-Ressel (2019), as shown in the proof of Appendix 2.B, and that $\kappa(t-s)$ is a decreasing L_2 -kernel, which is necessary in the following of the chapter to define the characteristic function of the log-cushion. Moreover, we can easily show that the initial forward variance curve at time $s = 0$ in the rough Heston model satisfies

$$\xi_0(t) = V_0 + (\theta - V_0) \frac{\lambda}{\nu} \int_0^t \kappa(s) ds. \quad (2.2.5)$$

2.3 CPPI in the rough Heston model

Under the assumption of frictionless trading for CPPI, we can first consider the stochastic volatility case of Prigent and Bertrand (2003) where the variance V_t of the underlying risky asset S_t in (2.2.1) is stochastic. They first show that the cushion value at any time $t \in [0, T]$ is

$$C_t = C_0 \exp \left\{ (r + m(\mu - r))t - \frac{1}{2}m^2 \int_0^t V_s ds + m \int_0^t \sqrt{V_s} dW_s \right\}. \quad (2.3.1)$$

Then, from the relation

$$S_t^m = S_0^m \exp \left\{ m \left(\mu t - \frac{1}{2} \int_0^t V_s ds \right) + m \int_0^t \sqrt{V_s} dW_s \right\}, \quad (2.3.2)$$

Prigent and Bertrand (2003) find that

$$C_t = S_t^m \alpha_t, \quad (2.3.3)$$

with $\alpha_t = (C_0/S_0^m) \exp\{\beta_t t\}$ and $\beta_t = -(m-1)r - \frac{1}{2}(m^2 - m) \frac{1}{t} \int_0^t V_s ds$. Finally, since the CPPI fund is the sum of the cushion and the floor, we get that $P_t = F_t + S_t^m \alpha_t$. On the contrary to the B&S framework, we no longer have that α_t is a constant but instead a random variable. Moreover, the portfolio value P_t now depends on the whole path of the variance process $\int_0^t V_s ds$ and hence, the short-memory of the rough Heston's variance process with $H < 1/2$ will have a strong

impact on P_t . However, expressions (2.3.2) and (2.3.3) will not be useful in our case, *i.e.* when the variance process exhibits a rough behavior as in equation (2.2.2). Indeed, as explained in Chapter 1, the rough Heston model is non-Markovian with respect to the current variance state V_s , which makes the quantity $\int_0^t V_s ds$ hard to compute and to obtain from Monte-Carlo simulations and even from more advanced numerical schemes such as the Hybrid scheme of Bennedsen et al. (2017). However, we can show that the rough Heston model is Markovian in the forward variance curve $\xi_s(t)$. Indeed, given the state vector $\xi_s(t)$, the dynamics of the model are well-determined. Hence, we will now capitalize on this Markov property with respect to $\xi_s(t)$ in order to compute more precisely and efficiently the value of the CPPI portfolio at any time $t \in [0, T]$. Denoting $X_t = \log C_t$, it is useful to rewrite equation (2.3.1) as

$$dX_t = (r + m(\mu - r)) dt - \frac{1}{2}m^2 V_t dt + m\sqrt{V_t} dW_t = (r + m(\mu - r)) dt + dY_t,$$

where

$$dY_t = -\frac{1}{2}m^2 V_t dt + m\sqrt{V_t} dW_t. \quad (2.3.4)$$

We now want to obtain the characteristic function of the forward variance model (Y, ξ) , *i.e.* of the process Y together with the family of processes $(\xi_s(t))_{t \geq s}$. From *Theorem 2.6.* of Gatheral and Keller-Ressel (2019) and from the dynamic of $\xi_s(t)$ given in equation (2.2.4), we have that that our forward variance model (Y, ξ) has an affine cumulant-generating function (CGF) in the sense that its CGF for $0 \leq s \leq T$ is of the form

$$\log \mathbb{E} \left[e^{u(Y_T - Y_s)} \middle| \mathcal{F}_s \right] = \int_s^T g(T - r, u) \xi_s(r) dr, \quad (2.3.5)$$

for all $\text{Re}(u) \in [0, 1]$ (or $u = iz$, $z \in \mathbb{R}$) and where $g(\cdot, u)$ is $\mathbb{R}_-$ -valued and continuous on $[0, T]$ for all $T > 0$ and $\text{Re}(u) \in [0, 1]$. This leads to the following theorem.

Theorem 2.1. *The function $g(\cdot, u) : \mathbb{R}_+ \rightarrow \mathbb{R}_-$ is the unique global continuous solution of the fractional Riccati equation² :*

$$g(t, u) = R_v \left(m, u, \int_0^t \kappa(t - r) g(r, u) dr \right) = R_v(m, u, (\kappa \star g)(t, u)), \quad (2.3.6)$$

where

$$R_v(m, u, w) = \frac{1}{2}m^2(u^2 - u) + m\rho uw + \frac{1}{2}w^2. \quad (2.3.7)$$

Proof. See Appendix 2.B. □

²The convolution operator $\star$ between two functions f, g is defined as : $(f \star g)(t) := \int_0^t f(t - r)g(r)dr$.

We now show how to determine the function $g(\cdot, u)$ for the CPPI under the rough Heston model (2.2.2). First let's denote

$$\psi(t, u) := \frac{1}{\nu} (\kappa \star g)(t, u) = \int_0^t (t-s)^{\alpha-1} E_{\alpha, \alpha}(-\lambda(t-s)) g(s, u) ds.$$

Thanks to Lemma A.2 of El Euch and Rosenbaum (2019), we find

$$D^\alpha \psi(t, u) + \lambda \psi(t, u) = g(t, u).$$

Then, using relations (2.3.6) and (2.3.7) with the fact that $(\kappa \star g)(t, u) = \nu \psi(t, u)$, we obtain

$$D^\alpha \psi(t, u) = \frac{1}{2} m^2 (u^2 - u) + (m\rho u\nu - \lambda) \psi(t, u) + \frac{1}{2} \nu^2 \psi(t, u)^2. \quad (2.3.8)$$

Finally, from the affine CGF (2.3.5) with $u = iz$ ($z \in \mathbb{R}$), we obtain the following characteristic function of (Y, ξ) in forward-variance form for the rough Heston model, conditionally on the initial state (Y_s, ξ_s) :

$$\Phi_s^Y(T, z) = \mathbb{E} [e^{izY_T} | Y_s, \xi_s] = \exp \left\{ izY_s + \int_s^T (D^\alpha \psi(T-r, iz) + \lambda \psi(T-r, iz)) \xi_s(r) dr \right\}. \quad (2.3.9)$$

From equation (2.2.5) and Appendix 2.A on fractional calculus, we can easily rewrite the initial characteristic function of (Y, ξ) at time $s = 0$ as

$$\Phi_0^Y(T, z) = \mathbb{E} [e^{izY_T} | Y_0, V_0] = \exp \left\{ izY_0 + \int_0^T \psi(T-r, iz) \left(\lambda\theta + \frac{V_0 r^{-\alpha}}{\Gamma(1-\alpha)} \right) dr \right\}. \quad (2.3.10)$$

The characteristic function of the log-cushion in forward variance form (X, ξ) at time $s = 0$ for the rough Heston model directly follows from (2.3.10) :

$$\begin{aligned} \Phi_0^X(T, z) = \mathbb{E} [e^{izX_T} | X_0, V_0] = \exp & \left\{ iz(X_0 + (r + m(\mu - r))T) \right. \\ & \left. + \int_0^T \psi(T-r, iz) \left(\lambda\theta + \frac{V_0 r^{-\alpha}}{\Gamma(1-\alpha)} \right) dr \right\}. \end{aligned} \quad (2.3.11)$$

The key issue we tackle in this chapter is to propose the most appropriate dynamic of the underlying risky asset under $\mathbb{P}$ using statistical methods on financial data such as to obtain the true properties of CPPI strategies. We are therefore interested in the dynamic of the portfolio under the real probability measure $\mathbb{P}$ and we hence need to estimate the parameters $(\alpha, \mu, \lambda, \theta, \nu, \rho)$ based on the observed time series of log-returns. Moreover, having a characteristic function in quasi-closed form for the log-cushion value X_T thanks to equation (2.3.11), we can obtain easily and efficiently several risk measures for C_T such as the VaR and TVaR, as well as its moments of all orders. This will be discussed in the following sections.

2.4 Estimation of the rough Heston model

We first choose as starting date $s = 0$, the 31st of July 2021. Since we decide throughout this chapter to consider a fixed investment horizon of $T = 1$ year for the CPPI strategy, we use the U.S. swap rate for this maturity equal to $r = 0.2546\%$. Then, in order to estimate the parameters of our model, we consider daily time series of log-returns from January 2017 to July 2021 for the S&P500 making a total of 1141 data points. In the literature so far, calibration of the rough Heston model has been exclusively done based on risk neutral option quotes such as in El Euch et al. (2019) or in Chapter 1. We now introduce a new approach providing a full econometric estimation of the rough Heston's parameters under the real probability measure $\mathbb{P}$, which is based on the Particle Markov Chain Monte Carlo (PMCMC) procedure of Andrieu et al. (2010). This algorithm is built from a sequential Monte-Carlo filter that we now describe.

2.4.1 Filtering

This section introduces the filtering technique that is used to determine the most likely evolution of the hidden variance process V . There is a considerable literature on the development of simulation-based methods to perform filtering of nonlinear Gaussian models. Leading contributions are reviewed in Doucet et al. (2001). We consider in particular the *Bayesian bootstrap filter* of Gordon et al. (1993), which is a Sequential Importance Resampling (SIR) filter in which the transition prior probability distribution is taken as importance function. This Bayesian bootstrap filter will be combined in the sequel with a Monte-Carlo Markov Chain to estimate the parameters of the rough Heston model but for the moment, we assume that its parameters are known. We consider a sample of discrete observations of log-returns $R_t = \ln \frac{S_{t+\Delta}}{S_t}$ where Δ is the time step between two observations. The sample is denoted by $r = \{r_1, r_2, \dots, r_n\}$ where $n = T/\Delta$. We denote v_j and r_j the realizations of $V_j = V_{t_j}$ and $R_j = R_{t_j}$ for $j = 1, \dots, n$ with $t_n = T$. The dynamic of log-returns is approached by applying Itô's Lemma on equation (2.2.1) and then discretizing the obtained equation, which gives

$$R_{j+1} = \left(\mu - \frac{V_j}{2} \right) \Delta + \sqrt{V_j} \Delta W_{j+1}.$$

The increment ΔW_j can be rewritten as $\Delta W_j = \sqrt{1-\rho^2} \Delta W_j^1 + \rho \Delta W_j^2$, where $\Delta W_j^1 \sim \mathcal{N}(0, \sqrt{\Delta})$ and independent from $\Delta W_j^2 \sim \mathcal{N}(0, \sqrt{\Delta})$. Therefore, we have

$$R_{j+1} = \left(\mu - \frac{V_j}{2} \right) \Delta + \sqrt{1-\rho^2} \sqrt{V_j} \Delta W_{j+1}^1 + \rho \sqrt{V_j} \Delta W_{j+1}^2. \quad (2.4.1)$$

The variance process (2.2.2) is discretized in the same manner by

$$V_{j+1} = V_0 + \sum_{k=0}^j \frac{(t_{j+1} - t_k)^{\alpha-1}}{\Gamma(\alpha)} \lambda (\theta - V_k) \Delta + \sum_{k=0}^j \frac{(t_{j+1} - t_k)^{\alpha-1}}{\Gamma(\alpha)} \nu \sqrt{V_k} \Delta W_{k+1}^2. \quad (2.4.2)$$

In practice, we prefer using the much more time-efficient simulation scheme of Abi Jaber (2019) for discretizing $(V_j)_{j=1,\dots,n}$. This author managed to build the rough variance process as superposition of infinitely many factors sharing the same Brownian motion but mean-reverting at different speeds. Limiting the number of factors allows to speed up the simulation scheme while still ensuring a extremely good proximity with the true original variance process, see Section 1.6 in Chapter 1. We denote the set of parameters $\vartheta = (\alpha, \mu, \lambda, \theta, \nu, \rho)$ and as mentioned above, it is assumed to be known for the moment. The particle filter estimates the most likely sample paths of the variance process. We denote by Δw_j^2 the realizations of ΔW_j^2 . The information on the realizations $V_j = v_j$ and $\Delta W_j^2 = \Delta w_j^2$ is stored in a *particle*, *i.e.* a vector denoted $p_j = (v_k, \Delta w_k^2)_{k=1:j}$. Contrary to classical stochastic volatility models, we now have to keep track of the whole sample path of the variance process and of the related Brownian increments in the particles from $k = 1$ to j due to the non-Markovian nature of the rough Heston model.

The visible information (log-returns) up to time t_j is stored in a vector $r_{1:j} = \{r_1, \dots, r_j\}$ and the density of the observation $f(r_j | p_j)$, conditionally to the information in the particle, is a Gaussian probability density function. More precisely, if we denote by $\phi(\cdot)$ the pdf of a $\mathcal{N}(0, 1)$, we have from equation (2.4.1) that

$$f(r_j | p_j) = \phi \left(\frac{r_j - \left(\mu - \frac{v_{j-1}}{2}\right) \Delta - \rho \sqrt{v_{j-1}} \Delta w_j^2}{\sqrt{1 - \rho^2} \sqrt{v_{j-1}} \sqrt{\Delta}} \right). \quad (2.4.3)$$

On the other hand, the transition density $f(p_{j+1} | p_j)$ is simulated based on equation (2.4.2). The initial pdf of p_0 is denoted $f(p_0)$. Then, using Bayesian arguments, one can show that the posterior distribution of p_j conditionally to the available information at t_j is equal to

$$f(p_j | r_{1:j}) = \frac{f(r_j | p_j)}{\int f(r_j | p_j) f(p_j | r_{1:j-1}) dp_j} f(p_j | r_{1:j-1}), \quad (2.4.4)$$

where we predict

$$f(p_j | r_{1:j-1}) = \int f(p_j | p_{j-1}) f(p_{j-1} | r_{1:j-1}) dp_{j-1}. \quad (2.4.5)$$

The calculation of $f(p_j | r_{1:j})$ can then be performed in three steps, according to Algorithm 1 in Appendix 2.C. First, in the prediction step, we approach $f(p_j | r_{1:j-1})$ by simulations, based on (2.4.5). In practice, the integral in (2.4.5) is replaced

by a Monte-Carlo simulation of N particles, denoted $p_j^{(i)}$ for $i = 1, \dots, N$. In the correction step, the density $f(p_j | r_{1:j})$ is computed via equation (2.4.4) and is replaced in simulations by the discrete probability of occurrence denoted $\beta_j^{(i)}$ for the particle $i = 1, \dots, N$. We hence have that $\beta_j^{(i)} = \mathbb{P}(p_j^{(i)} | r_{1:j})$, as given by equation (2.C.1). In the third step, we perform a resampling with replacement and with probabilities $\beta_j^{(i)}$, $i = 1, \dots, N$, in order to keep track of the most likely sample paths of the variance process. Finally, the filtered variance process for the period j is computed as

$$\widehat{V}_j = \mathbb{E}(V_j | r_{1:j}) = \sum_{i=1}^N v_j^{(i)} \beta_j^{(i)},$$

whereas the log-likelihood of the whole sample path is approached before resampling by the sum

$$\ln f(r | \vartheta) = \sum_{j=1}^n \ln f(r_j | r_{j-1}) = \sum_{j=1}^n \ln \left(\int f(r_j | p_j) f(p_j | r_{j-1}) dp_j \right) = \sum_{j=1}^n \ln \left(\sum_{i=1}^N \frac{1}{N} f(r_j | p_j^{(i)}) \right). \quad (2.4.6)$$

2.4.2 Econometric estimation

We have so far considered that the set of parameters ϑ is known, which is not the case in practice. An inherent problem of particle filters is that the estimate of the likelihood is not a smooth function of the parameters. Practically, model calibration by maximizing the log-likelihood (*e.g.* with gradient descent) is then uncertain. We here prefer using the PMCMC procedure which has been shown by Andrieu et al. (2010) and Johannes and Polson (2010) to be highly efficient for such financial estimation of stochastic volatility models. We now denote the set of unknown parameters by $\Theta = (\alpha, \mu, \lambda, \theta, \nu, \rho)$. We here again adopt a Bayesian approach and Θ is a multivariate random variable with realizations ϑ and with state space $\mathcal{X}$. The parameters posterior distribution, based on the observed sample $r = \{r_1, \dots, r_n\}$ is

$$f(\vartheta | r) = \frac{f(\vartheta) f(r | \vartheta)}{\int_{\mathcal{X}} f(\vartheta') f(r | \vartheta') d\vartheta'}, \quad (2.4.7)$$

where $f(\vartheta)$ and $f(r | \vartheta)$ denotes respectively the parameter prior distribution and the likelihood of the data. The density $f(\vartheta | r)$ is built by the PMCMC method. It generates samples from $f(\vartheta | r)$ by creating a Markov chain with the same distribution as the parameters' posterior one. Once the Markov chain has reached stationarity after a transient phase, called *burn-in* period, samples from the posterior distribution are simulated. Standard MCMC algorithm requires a pointwise estimate of $f(r | \vartheta)$, which is not available in our model. Instead, $f(r | \vartheta)$ is approached by its estimate (2.4.6), yield by the particle filter of the previous section.

The construction of the Markov chain consists of two steps, repeated iteratively for $k = 1, \dots, K$. At the beginning of the $(k + 1)^{th}$ iteration, we propose a candidate

parameter ϑ' from a proposal distribution $q(\vartheta' | \vartheta_k)$ given the previous state of the Markov chain, ϑ_k . The proposal distribution must have a support that covers the target distribution. In the second step, we determine if we update the state ϑ' . For this purpose, the acceptance probability of the Metropolis-Hasting algorithm described in Appendix 2.C (see Algorithm 2), is computed as follows

$$\rho(\vartheta', \vartheta_k) = \min \left\{ \frac{f(\vartheta' | r) q(\vartheta_k | \vartheta')}{f(\vartheta_k | r) q(\vartheta' | \vartheta_k)}, 1 \right\}. \quad (2.4.8)$$

This determines the probability that we assign the candidate parameter as the next state of the Markov chain, $\vartheta' \rightarrow \vartheta_{k+1}$. Intuitively, if we disregard the influence of the proposal distribution $q(\cdot, \cdot)$, a candidate is accepted if it increases the posterior likelihood $f(\vartheta' | r) > f(\vartheta_k | r)$. The presence of $q(\cdot, \cdot)$ in (2.4.8) allows a small decrease in the posterior likelihood, such as to explore the space of parameters χ .

In numerical applications, the transition distribution $q(\vartheta' | \vartheta_k)$ is assumed Gaussian, $\vartheta' \sim \mathcal{N}(\vartheta_k, \sigma_\vartheta I_m)$ with I_m the identity matrix. As this distribution is symmetric, we have $q(\vartheta_k | \vartheta') = q(\vartheta' | \vartheta_k)$ and the acceptance probability simplifies to

$$\rho(\vartheta', \vartheta_k) = \min \left\{ \frac{f(r | \vartheta') f(\vartheta')}{f(r | \vartheta_k) f(\vartheta_k)}, 1 \right\}.$$

The resulting samples of parameters $\vartheta_{1:K}$ (K iterations after a burn-in period) serves next to build the empirical distribution of $f(\vartheta | r)$, which is obtained by

$$f(\vartheta | r) \approx \frac{1}{K} \sum_{k=1}^K \delta_{\vartheta_k}(\vartheta),$$

where $\delta_{\vartheta_k}(\vartheta)$ are the Dirac atoms located at $\vartheta = \vartheta_k$, with equal weights. The estimate of parameters with respect to the posterior distribution is then

$$\hat{\Theta} = \mathbb{E}(\Theta | r) = \int_{\chi} f(\vartheta | r) \vartheta d\vartheta \approx \frac{1}{K} \sum_{k=1}^K \int_{\chi} \delta_{\vartheta_k}(\vartheta) \vartheta d\vartheta = \frac{1}{K} \sum_{k=1}^K \vartheta_k. \quad (2.4.9)$$

We now use the PMCMC algorithm to calibrate the set of parameters $\Theta = (\alpha, \mu, \lambda, \theta, \nu, \rho)$ of the rough Heston model (2.2.2). The standard deviation σ_ϑ of the transition distribution is set to $\sigma_\vartheta = (0.01\%, 0.01\%, 0.1\%, 0.01\%, 0.01\%, 0.1\%)$. Even if the procedure converges theoretically to parameter estimates, it is advisable to provide realistic initial parameters to improve the speed of convergence. Therefore, we impose $\vartheta_0 = (0.7, 0.05, 0.3, 0.04, 0.2, -0.7)$. The filter runs with $N = 1000$ particles and we perform $K = 20\,000$ iterations of the PMCMC procedure. Table 2.4.1 reports the rough Heston estimates with a burn-in period of 18 000 iterations. The obtained log-likelihood (2.4.6) with these parameters is equal to 3693.66.

Table 2.4.1: Parameter estimates of the rough Heston model for the S&P500 index and standard error (std. err.) of these estimates.

	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\nu}$	$\hat{\rho}$
Estimators PMCMC	0.8504	0.1446	0.2434	0.2122	0.3528	-0.5536
PMCMC std. err.	0.0128	0.0191	0.1272	0.0239	0.0183	0.0359

We observe that the estimate of the speed of mean-reversion λ is quite volatile with a high standard error. The other parameter estimates are more reliable. The advantage of our method based on a maximum likelihood estimator (MLE) for estimating α is the nice properties associated with such MLE. It is consistent, asymptotically unbiased and of minimal variance whereas the method of moments proposed by Gatheral et al. (2018) only exhibits consistency. Our estimator of α is hence more accurate. However, the method of moments of these authors based on the q -variation of the log-volatility can be used to confirm that the roughness index of the simulated rough Heston volatility with the MLE of Table 2.4.1 is consistent with the observed regularity of the past volatility time series. Using the methodology of these authors to estimate the smoothness of the S&P past volatility between January 2017 and July 2021, we find an estimated value of the Hurst index $\hat{H} = 0.2324$. We then simulate sample paths of our variance process V under the rough Heston model with $\alpha = 0.8504$ and $T = 4.5$ years (time length between January 2017 and July 2021), which gives $T \times 252 = 1142$ daily time steps. Based on these simulated sample paths, we find via the method of moments that the regularity of the corresponding volatility process is on average equal to $\hat{H} = 0.2505$, which is highly consistent with the value estimated based on the observed time series of volatility between 2017 and 2021.

2.5 Other models

In order to compare the rough Heston model with other models such as the Heston or B&S models in terms of CPPI properties, we need to derive the characteristic function of the log-cushion in these models. We show in this section that such log-cushion's characteristic function can be well defined analytically under affine Volterra processes and we can hence generalize the above results for this larger class of processes.

2.5.1 Heston model

First, the Heston model under the real probability measure $\mathbb{P}$ is given by

$$dS_t = S_t \mu dt + S_t \sqrt{V_t} dW_t, \quad (2.5.1)$$

$$dV_t = \lambda(\theta - V_t) dt + \nu \sqrt{V_t} d\widehat{W}_t. \quad (2.5.2)$$

The two Brownian motions $W, \widehat{W}$ are defined as above on $(\Omega, \mathcal{F}, \mathbb{P})$. We have that $V_0 = \sigma_0^2$ is the initial variance, θ is the long-run variance of V , λ denotes the speed of mean-reversion and ν is the volatility of variance parameter. From Bergomi and Guyon (2012) and Section 1.B.3, we know that the Heston model can be rewritten as a forward variance model with an affine CGF of the form of equation (2.3.5). Furthermore, based on *Example 2.10.* of Gatheral and Keller-Ressel (2019) and from Theorem 2.1, we obtain directly that the characteristic function of the log-cushion in the Heston model under $\mathbb{P}$ is

$$\begin{aligned} \Phi_0^X(T, z) = \mathbb{E} \left[e^{izX_T} \middle| X_0, V_0 \right] = \exp \left\{ iz \left(X_0 + (r + m(\mu - r))T \right) \right. \\ \left. + V_0 \psi(T, iz) + \lambda \theta \int_0^T \psi(r, iz) dr \right\}, \end{aligned} \quad (2.5.3)$$

and where $\psi(\cdot, iz)$ solves a (non-fractional) Riccati equation

$$\frac{\partial}{\partial t} \psi(t, iz) = -\frac{1}{2} m^2 (z^2 + iz) + (miz\rho\nu - \lambda) \psi(t, iz) + \frac{\nu^2}{2} \psi(t, iz)^2. \quad (2.5.4)$$

Note that this Riccati equation is consistent with the classical ODEs initially derived for the Heston model, such as in Gatheral (2011) (with $m = 1$). We also see that the fractional derivative of the rough Heston (which conveys the short memory to the model with $H < 1/2$) is replaced by a classical time derivative in the Riccati equation.

We again apply the PMCMC procedure above with the particle filter of Section 2.4.1 in order to estimate the parameters of the Heston model. The discretized variance process (2.4.2) now becomes under the Heston model

$$V_{j+1} = V_j + \lambda(\theta - V_j)\Delta + \nu\sqrt{V_j} \Delta W_{j+1}^2, \quad (2.5.5)$$

leading to the particle $p_j = (v_j, \Delta w_j^2)$ so that the process $(p_j)_{j=1, \dots, n}$ is a Markov chain (note that we do not need anymore to keep track of the whole sample path from $k = 1$ to j as in the rough Heston case). The set of unknown parameters becomes $\Theta = (\mu, \lambda, \theta, \nu, \rho)$ since we do not have to estimate the roughness parameter α , equal to 1 in the Heston model. We use the same standard deviation σ_ϑ and initial parameters ϑ_0 as in the rough Heston model (but removing the α -component). Finally, we again run the filter with $N = 1000$ particles and $K = 20\,000$ iterations and we obtain the following Table 2.5.1 with a log-likelihood equal to 3498.03. We first note a lower log-likelihood compared with the rough Heston model. This significant difference of log-likelihood confirms that the rough Heston model provides a better fit of the S&P log-returns from a statistical point of view. The average rate of return μ and the anti-correlation ρ are comparable between the two models. The mean reversion level θ is now lower compared to the rough Heston model while the speed of mean reversion λ is much higher. We again obtain a high standard error for the speed of mean-reversion level λ whereas the other estimates are more reliable.

Table 2.5.1: Parameter estimates of the Heston model for the S&P500 index and standard error of these estimates.

	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\nu}$	$\hat{\rho}$
Estimators PMCMC	0.1392	0.9256	0.1435	0.4831	-0.5506
PMCMC std. err.	0.0324	0.2656	0.0532	0.0106	0.0631

2.5.2 Affine Volterra processes

The methodology described above for deriving the characteristic function of the rough Heston and Heston models can be extended to the more general class of affine Volterra processes. This class of processes is defined in Abi Jaber et al. (2019) by a d -dimensional stochastic convolution equation of the form

$$Z_t = Z_0 + \int_0^t K_Z(t-s)b(Z_s)ds + \int_0^t K_Z(t-s)\sigma(Z_s)dW_s, \quad (2.5.6)$$

where W is a d -dimensional $\mathbb{P}$ -Brownian motion, where K_Z is a convolution kernel such that $K_Z \in L_{loc}^2(\mathbb{R}_+, \mathbb{R}^{d \times d})$ and where the coefficients b and σ satisfy the regularity and integrability conditions studied in Abi Jaber et al. (2019). This class of processes is called *affine* since the coefficient $a(z) = \sigma(z)\sigma(z)^\top$ (with $\sigma(z)$ positive semidefinite) and the coefficient $b(z)$ are both affine of the form

$$\begin{aligned} a(z) &= A^0 + z_1 A^1 + \cdots + z_d A^d, \\ b(z) &= b^0 + z_1 b^1 + \cdots + z_d b^d, \end{aligned}$$

for d -dimensional symmetric matrices A^i and column vectors b^i . We next introduce the $d \times d$ matrix

$$B = (b^1 \quad \cdots \quad b^d),$$

and for any row vector $u \in \mathbb{C}^d$, we define the row vector

$$A(u) = (uA^1 u^\top, \dots, uA^d u^\top). \quad (2.5.7)$$

We then recall from the log-cushion dynamic (2.3.4) that

$$Y_t = Y_0 - \frac{1}{2}m^2 \int_0^t V_s ds + m \int_0^t \sqrt{V_s} dW_s. \quad (2.5.8)$$

Moreover, we now consider that the variance process V follows the general (one-dimensional) convolution equation given by

$$V_t = V_0 + \int_0^t K(t-s)(\beta + bV_s)ds + \int_0^t K(t-s)(\alpha + a\sqrt{V_s})d\widehat{W}_s, \quad (2.5.9)$$

again with a constant correlation ρ such that $\langle W, \widehat{W} \rangle_t = \rho t$, with $\beta, b, \alpha, a \in \mathbb{R}$ and a convolution kernel $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ such that :

K is strictly positive and completely monotone. There is $\gamma \in (0, 2]$ such that

$$\int_0^h K(t)^2 dt = O(h^\gamma) \text{ and } \int_0^T (K(t+h) - K(t))^2 dt = O(h^\gamma) \text{ for } T < \infty. \quad (2.5.10)$$

Hence, we take $d = 2$ and consider the two-dimensional process $Z = (Y, V)^\top$. In order to have an affine Volterra process Z compatible with the given equation (2.5.8) of Y , we need to impose the coefficient $a(z)$ of Z to be of the form

$$A^0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad A^1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad A^2 = \begin{pmatrix} m^2 & m\rho a \\ m\rho a & a^2 \end{pmatrix}. \quad (2.5.11)$$

This amounts to set $\alpha = 0$ in the variance process 2.5.9. Otherwise, there does not exist a positive semidefinite matrix $\sigma(z)$ which is such that $a(z) = \sigma(z)\sigma(z)^\top$ is affine. The coefficients $b(z)$ and the kernel of the two-dimensional affine Volterra process Z are given by

$$b^0 = \begin{pmatrix} 0 \\ \beta \end{pmatrix} \quad B = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & b \end{pmatrix} \quad K_Z = \begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix}. \quad (2.5.12)$$

We clearly see that the Heston model described in Section 2.5.1 can be retrieved by setting the kernel $K \equiv 1$ and the coefficients $a = \nu$, $\beta = \lambda\theta$ and $b = -\lambda$ in the variance process. For the rough Heston model, we need to impose a power law kernel $K(t) = t^{\alpha-1}/\Gamma(\alpha)$ for the variance process instead, which is known to satisfy the condition (2.5.10). We now derive from Abi Jaber et al. (2019) the characteristic function of the log-cushion under the two-dimensional affine Volterra process Z for general coefficients a, β, b and for general convolution kernel K .

Theorem 2.2. *Let $Z = (Y, V)^\top$ be a solution of (2.5.6) with $K_Z \in L^2(\mathbb{R}_+, \mathbb{R}^{2 \times 2})$, where we assume that K satisfies (2.5.10) and where $A(\cdot)$ is given by (2.5.7). Let $u = (u_1, u_2) \in \mathbb{C}^2$ with $\text{Re } u_1 \in [0, 1]$ and $\text{Re } u_2 \leq 0$ and assume that $\psi = (\psi_1, \psi_2)^\top \in L^2([0, T], \mathbb{C}^2)$ solves the Riccati-Volterra equation*

$$\psi = uK_Z + \left(\psi B + \frac{1}{2} A(\psi) \right) \star K_Z. \quad (2.5.13)$$

Using the form of K_Z and of the coefficients $b(z)$ and $a(z)$ given by (2.5.11) and (2.5.12), we obtain

$$\psi_1 = u_1, \quad (2.5.14)$$

$$\psi_2 = u_2 K + K \star \mathcal{R}_\Psi(\psi), \quad (2.5.15)$$

where we define

$$\mathcal{R}_\Psi(\psi) = \frac{1}{2} m^2 (\psi_1^2 - \psi_1) + (b + m\psi_1 \rho a) \psi_2 + \frac{a^2}{2} \psi_2^2 \quad \text{and} \quad \mathcal{R}_\Phi(\psi_2) = \beta \psi_2.$$

Then, the Fourier-Laplace transform of Z_T is given by

$$\mathbb{E}[e^{uZ_T}] = e^{\phi(T) + \chi(T)V_0 + u_1 Y_0}, \quad (2.5.16)$$

where

$$\phi' = \frac{d}{dt}\phi = \mathcal{R}_\Phi(\psi_2), \quad \chi' = \frac{d}{dt}\chi = \mathcal{R}_\Psi(\psi),$$

and

$$\phi(0) = 0, \quad \chi(0) = u_2.$$

Proof. Cfr Appendix 2.D.

However, we are especially interested in the characteristic function of the log-cushion X which can be obtained from the dynamic of Y using equation (2.3.4). Therefore, we now study the following affine Volterra process $\tilde{Z} = (X, V)^\top$ and obtain the next Corollary 2.3 with the same assumptions as in the previous theorem.

Corollary 2.3. *The Fourier-Laplace transform of the the log-cushion X_T at maturity T associated with the affine Volterra process $\tilde{Z}$ is given by*

$$\Phi_0^X(T, u_1) = \mathbb{E}[e^{u_1 X_T}] = e^{\phi(T) + \chi(T)V_0 + u_1(X_0 + (r + m(\mu - r))T)}, \quad (2.5.17)$$

with ϕ and χ defined by

$$\phi' = \frac{d}{dt}\phi = \mathcal{R}_\Phi(\psi_2), \quad \chi' = \frac{d}{dt}\chi = \mathcal{R}_\Psi(\psi),$$

where $\mathcal{R}_\Phi(\psi_2)$ and $\mathcal{R}_\Psi(\psi)$ are defined as in Theorem 2.2 and with initial conditions given by

$$\phi(0) = 0, \quad \chi(0) = 0.$$

The process $\psi = (\psi_1, \psi_2)^\top \in L^2([0, T], \mathbb{C}^2)$ is now the solution of the following Riccati-Volterra equation

$$\psi_1 = u_1, \quad (2.5.18)$$

$$\psi_2 = K \star \mathcal{R}_\Psi(\psi) = K \star \left(\frac{1}{2}m^2(u_1^2 - u_1) + (b + mu_1\rho a)\psi_2 + \frac{a^2}{2}\psi_2^2 \right). \quad (2.5.19)$$

Proof. Cfr Appendix 2.E.

It is important to note that we only consider here the unconditional characteristic function of affine Volterra processes at time $s = 0$. Indeed, Theorem 2.2 and

Corollary 2.3 do not hold anymore at a later time s since we cannot write the conditional characteristic function $\mathbb{E}[e^{u_1 X_T} | \mathcal{F}_s]$ as a linear function of $\phi(\cdot)$ and $\chi(\cdot)$. Based on this Corollary 2.3, we can easily recover the characteristic function of the log-cushion under the Heston model, which is given by equations (2.5.3) and (2.5.4) and under the rough Heston model, given by (2.3.11) and (2.3.8). It suffices to take $u_1 = iz$, $a = \nu$, $\beta = \lambda\theta$ and $b = -\lambda$ with $K \equiv 1$ as convolution kernel in the Heston model and $K(t) = t^{\alpha-1}/\Gamma(\alpha)$ in the rough Heston model. Moreover, from Ackermann et al. (2020), we note that we could extend the results above by considering a time-dependent mean-reversion level $\theta(t)$ instead of the constant θ , and hence work with the generalized rough Heston model of El Euch and Rosenbaum (2019) as in Chapter 1, equation (1.3.6).

We finally conclude that the class of Volterra processes tends to be suitable for building CPPI strategies that are analytically tractable through their characteristic function. Such affine Volterra processes allow to easily derive CPPI properties under various market conditions. Indeed, recent empirical studies suggest that the memory properties of market fears are unstable and change over time, leading to fluctuating behavior of financial markets as argued by Caporale et al. (2018). Following Cadoni et al. (2017), roughness is also associated with risks in pension fund portfolios where such CPPI strategies are often used. Affine Volterra processes can also be useful for studying CPPI strategies based on underlying assets related to the commodity market, see Benth et al. (2020) or Alfeus et al. (2024).

2.5.3 Black and Scholes model

Finally, for the sake of comparison with what has already been done in the literature, we analyze the characteristic function of the log-cushion in the B&S model. In this B&S model, we have that the log-cushion dynamic can be rewritten, in a similar way as in equation (2.3.4), by

$$dX_t = (r + m(\mu - r))dt - \frac{1}{2}m^2\sigma^2 dt + m\sigma dW_t. \quad (2.5.20)$$

Hence, it directly follows that the characteristic function of the log-cushion is given by

$$\begin{aligned} \Phi_0^X(T, z) = \mathbb{E}[e^{izX_T} | X_0] = \exp \left\{ iz \left(X_0 + T(r + m(\mu - r)) \right) \right. \\ \left. - \frac{1}{2}m^2(z^2 + iz)\sigma^2 T \right\}. \end{aligned} \quad (2.5.21)$$

There are no latent variables in this model which implies that we do not need to use a particle filter and that we can directly apply the Metropolis-Hasting algorithm with unknown parameters $\Theta = (\mu, \sigma)$. Since the log-returns follow $R_j \sim \mathcal{N}((\mu - \sigma^2/2)\Delta, \sigma^2\Delta)$ for $j = 1, \dots, n$, the log-likelihood of the sample

is now directly obtained by

$$\ln f(r|\vartheta) = -\frac{n}{2} \ln(2\pi\sigma^2\Delta) - \ln \left(\frac{\sum_{j=1}^n \left(r_j - \left(\mu - \frac{\sigma^2}{2} \right) \Delta \right)^2}{2\sigma^2\Delta} \right).$$

Again with $K = 20\,000$ iterations, we find a log-likelihood of 3380.25 and the following parameter estimates.

Table 2.5.2: Parameter estimates of the B&S model for the S&P500 index (01/2017-06/2021) and standard error of these estimates.

	$\hat{\mu}$	$\hat{\sigma}$
Estimators PMCMC	0.1099	0.1988
PMCMC std. err.	0.0657	0.0042

The log-likelihood is the lowest of the analyzed models, leading to the poorest statistical fit. We obtain a very precise estimator of the volatility σ , which is lower compared to the mean volatility levels derived in the two previous models. The mean rate of return μ is also lower in the B&S model and has a quite high standard error.

2.5.4 (Fractional) Riccati equations

In order to obtain the characteristic function of the log-cushion in the rough Heston model, we need to solve the fractional Riccati equation (2.3.8). Based on Momani and Shawagfeh (2006), we use an Adomian decomposition method combined with Padé approximants to obtain numerically the solution of such fractional Riccati equations. This method is highly effective and stable in order to capture the essential behavior of the solution $\psi(t, u)$ in the fractional Riccati case and provides solutions that are consistent with the numerical scheme described in the classical paper of Diethelm et al. (2002) for finding solutions of fractional differential equations. For the Heston model, we need to solve the ordinary Riccati equation (2.5.4). No analytical solution can be found to this first-order nonlinear ODE and we hence also resort to the Adomian method of Momani and Shawagfeh (2006) with $\alpha = 1$ to derive numerically the solution $\psi(t, u)$.

We now show how to compute the moments of different orders for the portfolio value $P_T = C_T + F_T$ at maturity. We mainly consider the Heston and rough Heston models, which are two important members of our class of affine Volterra processes $\tilde{Z}$. We also compare these results to the B&S framework for the log-cushion dynamic.

2.6 Moments

From the characteristic function of the log-cushion, the moments of order k of the cushion C_T can be directly obtained by imposing $iz = k$. Indeed, we have

$$\Phi_0^X(T, -ki) = \mathbb{E} [e^{kX_T}] = \mathbb{E} [e^{k \log C_T}] = \mathbb{E} [C_T^k]. \quad (2.6.1)$$

Based on the estimated parameters (2.5.1) of the Heston and (2.4.1) of the rough Heston, it can be shown for $k \leq 4$ using Theorem 3 of Gerhold et al. (2019) that $\sup \{t \geq 0 : \mathbb{E}[e^{kX_t}] < \infty\} = \infty$ and hence, that the first four moments of the cushion are finite for any time horizon. We consider here the situation at time $s = 0$ where the initial amount in the fund $P_0 = 100$ and where the guaranteed amount $G(= F_T)$ at maturity $T = 1$ is 95. We find that $F_0 = 95e^{-r} = 94.77$ and hence that $C_0 = P_0 - F_0 = 5.23$.

For the multiple m , applying the extreme value approach of Bertrand and Prigent (2016), we are able to provide an empirical upper bound of m based on our dataset of past log-returns for the S&P500. Considering a management period of 1 month (21 days, $n = 12$) and a risk tolerance ε of 0.5%, we find empirically an upper bound $m \leq 4.1117$ such that the shortfall probability satisfies $P^{SF} := \mathbb{P}[\exists j \in [1, n] : C_{t_j} \leq 0] \leq \varepsilon$ by following the methodology of these authors. We therefore choose a value $m = 4$ satisfying this upper bound such that the worst daily return anticipated by the fund manager over the investment horizon is 25%. Using the models introduced above, we are able to find the k^{th} order moment of the cushion value C_T and we hence obtain the following Table 2.6.1 describing the expectation, standard deviation, skewness and kurtosis of the portfolio value P_T at maturity, using the relation $\mathbb{E}[P_T^k] = \mathbb{E}[(C_T + F_T)^k]$.

Table 2.6.1: Comparison of the first four moments of P_T based on the characteristic function of the log-cushion under each of the three estimated models.

	Expectation	Std. dev.	Relative Skew.	Relative Kurt.
B&S	103.0599	7.5695	3.6458	33.5037
Heston	104.0626	8.1671	2.7833	55.3041
Rough Heston	104.2598	6.2334	2.1259	68.7292

First, we have logically that the expected value of the fund is lower in the B&S model due to the lower rate of return μ . We also see that standard deviation and skewness are lower in the rough Heston model due to the much lower speed of mean reversion λ . However, the main difference comes from the 4^{th} moment which is higher in the rough Heston case and lower in the B&S model. These differences can be explained by the absence of leverage effect in the B&S model and by the roughness property of the variance process under the rough Heston case, as we will

see below with Table 2.6.2. Note that moments of the fund obtained via simulations such as in Prigent and Bertrand (2003) are consistent with the values derived from our method based on the characteristic function of the log-cushion. Moreover, the methodology derived in this Chapter 2 overcomes the instability and slowness of these simulations and allows to accurately and efficiently obtain the moments of the fund value P_T .

Finally, we want to emphasize the impact of rough volatility on the portfolio's moments. We therefore impose for the rough Heston model the same parameters $\mu, \lambda, \theta, \nu, \rho$ as in the Heston model (see Table 2.5.1) and we consider for the roughness index α a value of 0.75. For the B&S model, we also take $\mu = 0.1392$ as in the Heston model and σ equal to the average Heston volatility $\frac{1}{T} \int_0^T \mathbb{E}[\sqrt{V_s}] ds = 0.2156$. Using these comparable parameters, we obtain the following Table 2.6.2.

Table 2.6.2: Comparison of the first four moments of P_T based on the characteristic function of the log-cushion under each of the three models with comparable parameters.

	Expectation	Std. dev.	Relative Skew.	Relative Kurt.
B&S	104.0626	9.5189	4.3010	48.4466
Heston	104.0626	8.1671	2.7833	55.3041
Rough Heston	104.0626	8.1870	2.7220	116.7027

First, we have that the B&S model provides a standard deviation and a skewness that are higher than in the other models. This is due to the negative correlation ρ of the Heston and rough Heston models which decreases their moment values. Therefore, since the B&S model does not take into account the leverage effect, it tends to overestimate the moments of the fund P_T . Concerning the kurtosis, the absence of heavy tails in the B&S model compensates for the absence of leverage effect and we hence obtain a lower 4th moment than with the two other models. Next, the main difference between the Heston and rough Heston models comes from the 4th moment, with a much higher kurtosis in the rough Heston case. The roughness of the volatility hence strongly impacts the tail of the portfolio distribution and the occurrence probability of extreme events affecting the CPPI value. Indeed, as explained in Paulot and Lacroze (2009), CPPI is very sensitive to tail events and its value distribution mainly depends on the mean volatility and on the left tail of the underlying distribution. Hence, this rough behavior of the underlying's volatility is crucial and must be taken into account in order to better model the tails of the CPPI distribution and assess the true CPPI properties.

In a similar way as for the calculation of the moments of the portfolio value P_T , we now show that we can derive efficiently and consistently several risk measures as well as the density of P_T based on the characteristic function of the log-cushion. These results will be obtained both with the true estimated parameters of each model, as well as with the comparable parameters based on the Heston model as

in Table 2.6.2.

2.7 Risk Measure and probability density

Based on Dhaene et al. (2006), we define the Value-at-Risk (VaR) in this chapter as the ε -quantile of the portfolio value at the chosen maturity T . This risk measure is used to appraise the shortfall risk and despite that it is not a coherent risk measure (see Artzner et al. (1999)), the Value-at-Risk still remains an essential tool for asset allocation. If $\varepsilon \in (0, 1)$ is the minimum probability level accepted by the investor, then the VaR is defined as the shortfall level, q , such that

$$\begin{aligned} VaR_\varepsilon(P_T) &:= \inf \{q \in \mathbb{R} \mid \mathbb{P}(P_T \leq q) \geq \varepsilon\} = \inf \{q \in \mathbb{R} \mid \mathbb{P}(C_T \leq q - F_T) \geq \varepsilon\} \\ &= \inf \{q \in \mathbb{R} \mid \mathbb{P}(X_T \leq \log(q - F_T)) \geq \varepsilon\} \end{aligned} \quad (2.7.1)$$

where $X_T = \log C_T$. If we then denote the probability density function of X_T by $f_T(\cdot)$, we then have

$$VaR_\varepsilon(P_T) = \inf \left\{ q \in \mathbb{R} \mid \int_{-\infty}^{\log(q - F_0 e^{rT})} f_T(x) dx \geq \varepsilon \right\}. \quad (2.7.2)$$

Another common measure of risk is the Tail Value-at-Risk (TVaR), which is the expected shortfall level if the portfolio value breaches $VaR_\varepsilon(P_T)$, defined as

$$\begin{aligned} TVaR_\varepsilon(P_T) &:= \mathbb{E}[P_T \mid P_T \leq VaR_\varepsilon(P_T)] \\ &= \mathbb{E}[C_T + F_0 e^{rT} \mid C_T \leq VaR_\varepsilon(P_T) - F_0 e^{rT}] \\ &= \mathbb{E}[e^{X_T} \mid X_T \leq \log(VaR_\varepsilon(P_T) - F_0 e^{rT})] + F_0 e^{rT} \end{aligned} \quad (2.7.3)$$

since we are interested for this chapter in the left-tail of the portfolio distribution. Equation (2.7.3) can also be rewritten in terms of the pdf $f_T(\cdot)$ as follows

$$TVaR_\varepsilon(P_T) = \frac{1}{\varepsilon} \int_{-\infty}^{\log(VaR_\varepsilon(P_T) - F_0 e^{rT})} e^x f_T(x) dx + F_0 e^{rT}. \quad (2.7.4)$$

The TVaR is a coherent risk measure, also called expected shortfall and can identically be defined as the conditional tail expectation when applied to continuous random variables.

The probability distribution of the log-cushion X_T (conditionally to X_0 and V_0) under the above models is not analytically determined (except for the B&S model). However, its characteristic function $\Phi_0^X(T, z)$, which is also the Fourier Transform of its density, has a quasi-closed form under these models. By inverting numerically the characteristic function of X_T , it is hence possible to determine its density of probability and to compute numerically the VaR and TVaR by equations (2.7.2) and (2.7.4). This numerical scheme, adapted from Hainaut (2011), relies on the Fast Fourier Transform (FFT) algorithm and is detailed in Appendix 2.F.

2.7.1 Application

2.7.1.1 Comparable parameters

The log-cushion density obtained numerically from the FFT algorithm (see Appendix 2.F) is first plotted below in Figure 2.7.1 for each of the studied models based on the Heston parameters (as used in Table 2.6.2). The number of steps involved in the FFT algorithm is set to $M = 2048$.

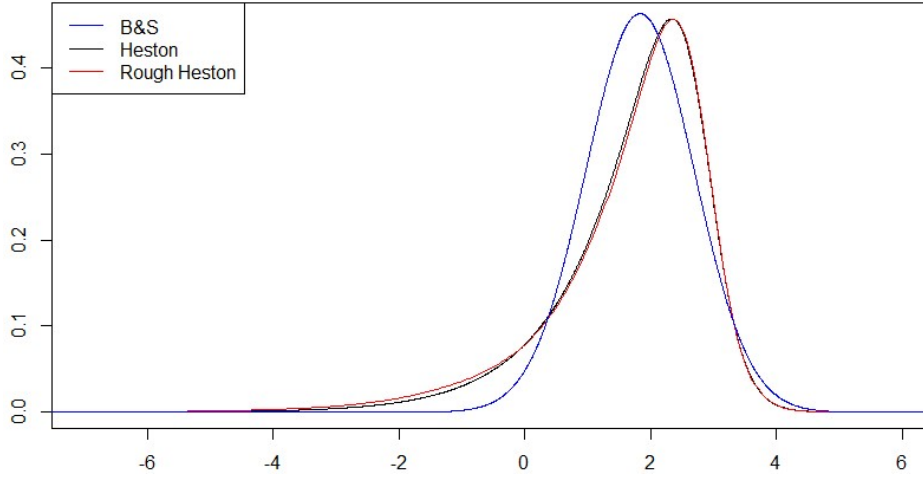


Figure 2.7.1: Density of the log-cushion X_T at maturity $T = 1$ for the three models with comparable parameters and $m = 4$.

We first observe that the log-cushion is normally distributed under the B&S model, which explains the shape of the blue curve in Figure 2.7.1. Compared with the other models, we see that the probability density function of the B&S log-cushion is shifted to the right with more weight attributed to high values of the log-cushion. We also see that log-cushion values are more dispersed around the mean in the B&S model. This explains the higher volatility of the portfolio value P_T in Table 2.6.2.

The density functions of the log-cushion under the Heston and rough Heston models are extremely similar. Both are negatively-skewed with fat left tails. First, this negative skewness of the log-cushion explains why the skewness of the portfolio value P_T in Table 2.6.2 is lower than in the B&S model when taking the exponential of the distribution (but still positive since the effect of the exponential function on the skewness overcomes the negative skewness of the log-cushion). Secondly, the fatter left tail observed for the black and red curves below is the reason why we obtain a higher kurtosis in Table 2.6.2, compared with the B&S model. Finally, based again on Table 2.6.2, we also observe that this is logical to obtain very similar densities between the Heston and rough Heston models since the three first

moments are extremely close to each other. Only the kurtosis differs between these two models, which is indeed confirmed on Figure 2.7.1 with a fatter left tail for the rough Heston in red compared with the Heston curve in black. A higher multiplier value m would have emphasized these effects.

Based on the probability density function of the log-cushion, we can now derive its quantiles and hence use formula (2.F.3) to compute the VaR of the portfolio value P_T . We then obtain the following Table 2.7.1 describing five percentiles of interest for P_T (mainly describing the left tail of the distribution) with $m = 4$.

Table 2.7.1: Comparison of $VaR_\varepsilon(P_T)$ for different levels ε between models (comparable parameters) at $T = 1$.

	0.5%	1%	5%	25%	50%
B&S	95.6766	95.8388	96.5130	98.4903	101.2466
Heston	95.0566	95.1140	95.5953	98.2514	102.1060
Rough Heston	95.0272	95.0674	95.4746	98.0959	101.9930

This Table 2.7.1 confirms and details Figure 2.7.1 of the log-cushion density. The rough Heston model tends to have the lowest values for small percentiles (and hence be the most risky) since it has the fattest left tail, closely followed by the Heston model. We also clearly see that the B&S model exhibits the highest values for small quantiles, which is again explained by its lower kurtosis. The trend reverses for higher quantiles such as the median $\varepsilon = 50\%$, where we obtain larger values for the Heston and rough Heston models compared with the B&S model. We now turn to the TVaR of the portfolio value P_T which is defined by formula (2.7.4), and we obtain the following Table 2.7.2 for all our models with $m = 4$.

Table 2.7.2: Comparison of $TVaR_\varepsilon(P_T)$ at levels 0.5%, 1%, 5% and 25% between models (comparable parameters) for $T = 1$.

	0.5%	1%	5%	25%
B&S	95.5316	95.6478	96.1108	97.2603
Heston	95.0284	95.0569	95.2950	96.5821
Rough Heston	95.0120	95.0310	95.2182	96.4484

By definition, we have that the TVaR is below the VaR at corresponding levels. Again, we confirm that the rough Heston model has the fattest left tail and is the most risky with significant probability attributed to very small values of P_T , while the B&S portfolio value exhibits the highest TVaR for small percentile levels. Finally, it can again be confirmed that the values of VaR and TVaR derived in Table 2.7.1 and Table 2.7.2 are consistent and close to the values obtained from simulations. However, simulation-based methods (Monte-Carlo, Hybrid scheme,

etc.) are way more slow and less stable than working directly with the characteristic function of the log-cushion, especially for the rough Heston model due to the non-Markovian nature of its variance process.

2.7.1.2 True estimated parameters

We now derive the log-cushion density (Figure 2.7.2) as well as the VaR and TVaR of the portfolio value P_T (Tables 2.7.3 and 2.7.4), based on the original parameters of each model given in Table 2.6.1 since these parameters are those of true interest.

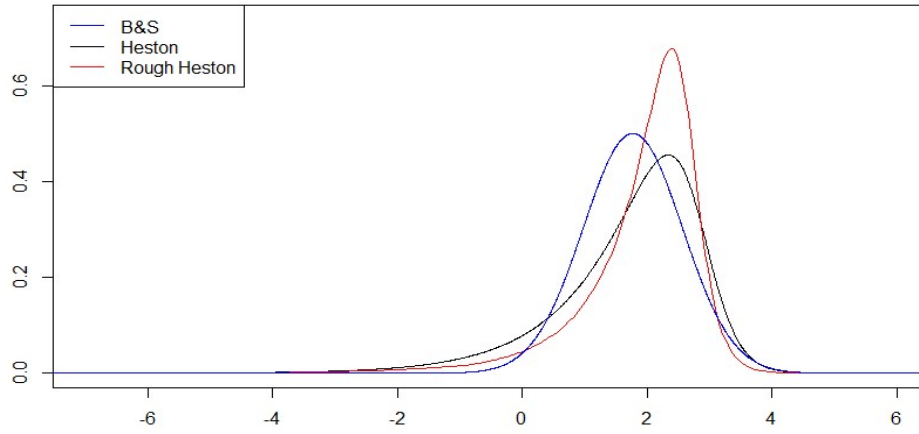


Figure 2.7.2: Density of the log-cushion X_T at maturity $T = 1$ for each model with the true estimated parameters and $m = 4$.

Table 2.7.3: Comparison of $VaR_\epsilon(P_T)$ at levels 0.5%, 1%, 5%, 25% and 50% between the three models with the true estimated parameters and $T = 1$.

	0.5%	1%	5%	25%	50%
B&S	95.7578	95.9212	96.5856	98.4303	100.8681
Heston	95.0566	95.1140	95.5953	98.2514	102.1060
Rough Heston	95.1368	95.2596	96.1420	99.7313	103.4551

Despite having the highest kurtosis in Table 2.6.1, the rough Heston model with the true estimated parameters is less risky than the Heston model with higher risk measures in the rough Heston (VaR, TVaR) for small quantiles. Indeed, this can be explained by the fact that the rough Heston model has a substantially lower standard deviation and skewness in Table 2.6.1 compared with the Heston model, which also impacts the riskiness and gap risk of the CPPI strategy (as we will see in

Table 2.7.4: Comparison of $TVaR_\varepsilon(P_T)$ at levels 0.5%, 1%, 5% and 25% between models with the true estimated parameters and $T = 1$.

	0.5%	1%	5%	25%
B&S	95.6096	95.7222	96.1842	97.2726
Heston	95.0284	95.0569	95.2950	96.5821
Rough Heston	95.0699	95.1320	95.5921	97.6052

the next section). The impact of the kurtosis on these two models can however be seen in Tables 2.7.3 and 2.7.4 with lower differences between risk measures at small levels (0.5%, 1%, 5%) and higher differences at larger levels (25%, 50%). Figure 2.7.2 confirms these findings.

2.8 Discrete-time CPPI

Since we have assumed continuous trading on the whole investment horizon $[0, T]$, we cannot take into account the gap risk, *i.e.* the risk of violating the floor protection. We can indeed observe in Table 2.7.1 that the final value of the portfolio P_T is systematically above the guarantee $G = 95$, even for very small quantiles. This gap risk is however caused by liquidity constraints (see the Introduction) as well as steep drops in the risky asset occurring before the fund manager can rebalance the portfolio and it therefore needs to be managed. Such gap risk is discussed thoroughly in Cont and Tankov (2009) and in Bertrand and Prigent (2016). Nevertheless, both liquidity constraints and gap risk can be modeled in the setup of Balder et al. (2009) where the price dynamic of the risky asset S is again described by a continuous-time stochastic process but where trading is now restricted to discrete times. In their framework, they consider the sequence τ^n of equidistant refinements of the interval $[0, T]$, *i.e.* $\tau^n = \{t_0^n = 0 < t_1^n < \dots < t_{n-1}^n < t_n^n = T\}$. This sequence represents the effective trading times and as above, we consider a management period of one month so that $n = 12$ (since $T = 1$). Following Balder et al. (2009), the self-financing condition of the CPPI portfolio implies the following recursive relation for the cushion C with $k = 1, \dots, n - 1$:

$$C_{t_{k+1}^n} = \begin{cases} C_{t_k^n} \left(m \frac{S_{t_{k+1}^n}}{S_{t_k^n}} - (m - 1)e^{rT/n} \right) & \text{if } C_{t_k^n} > 0 \\ C_{t_k^n} e^{rT/n} & \text{if } C_{t_k^n} \leq 0. \end{cases} \quad (2.8.1)$$

Balder et al. (2009) consider that the dynamic of the price S simply follows a geometric Brownian motion and hence, they manage to derive closed-form formula for the shortfall probability $P^{SF} = \mathbb{P}(\exists k \in [1, n] : C_{t_k^n} \leq 0)$, for the expected shortfall $ESF = \mathbb{E}[-C_{t_n^n} | \exists k \in [1, n] : C_{t_k^n} \leq 0]$ and for the gap risk $GPR = \mathbb{E}[-C_{t_n^n} \mathbb{1}_{\{\exists k \in [1, n] : C_{t_k^n} \leq 0\}}]$. In order to handle and compute these gap risk measures associated with our Heston and rough Heston models, we need to

simulate the price process in (2.8.1) under both models since no closed-form formula exist for P^{SF} , ESF and GPR in such cases. We use the same simulation techniques as described in Chapter 1, Section 1.6. For the rough Heston, we hence again consider the time-efficient simulation scheme of Abi Jaber (2019) based on a lifted version of the Heston model, which works well in practice. We first obtain Table 2.8.1 with the comparable parameter estimates derived in Table 2.6.2.

Table 2.8.1: Shortfall probability, Expected shortfall and Gap risk for the three models with comparable parameters.

	P^{SF}	ESF	Gap risk
B&S	1.08×10^{-5}	0.24237	2.62×10^{-6}
Heston	0.0228	0.5100	0.0116
Rough Heston	0.0438	0.8125	0.0356

We clearly see that the rough Heston model provides the highest risk measures due to its fatter left tail. The B&S CPPI strongly underestimates the gap risk and gives measures way below the two other models. Moreover, we have the following Table 2.8.2 giving these different risk measures based on the true estimated parameters.

Table 2.8.2: Shortfall probability, Expected shortfall and Gap risk for the three models with $m = 4$ and the true estimated parameters.

	P^{SF}	ESF	Gap risk
B&S	1.65×10^{-6}	0.1990	3.28×10^{-7}
Heston	0.0228	0.5100	0.0116
Rough Heston	0.0053	0.4373	0.0025

What is fundamental to note is that the rough Heston model with the true estimated parameters provides a shortfall probability in accordance with the empirical one associated with a multiple $m = 4$. Indeed, recall from Section 2.6 that the estimated shortfall probability ε based on the empirical log-return data with $m = 4.1117$ is equal to 0.5%. Hence, we see that the B&S model strongly underestimates the shortfall probability of the CPPI strategy while the Heston model overestimates it and tends to be too risky. The fact that the rough Heston model manages to remarkably reproduce such empirical shortfall probability means that rough volatility models indeed better reproduce tail events and are more consistent with the gap risk derived from empirical log-return data. This again confirms the importance of taking into account the roughness of volatility when considering and assessing CPPI strategies. Note that the client of a CPPI fund may be insured against this gap risk in such a way that he can recover from the bank the loss made if the fund breaks the floor. In this case, the share of a CPPI fund is a

financial product with an optional component, *i.e.* the sum of a self-financing portfolio corresponding to the uninsured CPPI strategy and a put option with payoff $-C_T \mathbb{1}_{\{\exists k \in [1, n] : C_{t_k} \leq 0\}}$. Practically, our methodology allows to better assess these gap and liquidity risks since it provides a better modeling of the tails of the CPPI distribution and since it is more consistent with financial data. Clients and practitioners can hence better hedge against these risks. Moreover, transaction costs could also be taken into consideration with this discrete setting, see Balder et al. (2009).

2.9 Conclusion

This chapter proposes to derive and study the properties of CPPI strategies based on the characteristic function of the log-cushion. The class of affine Volterra processes appears to be particularly adapted for building such CPPI strategies in frictionless markets since their log-cushion's characteristic function is given in quasi-closed form and can be evaluated rapidly using numerical schemes. This method hence generalizes the results previously established in the literature for a large class of models, especially when the variance process associated with the risky asset is non-Markovian in the current variance state. Through the choice of the coefficients a and b as well as the convolution kernel K of affine Volterra processes, various and more realistic dynamics can be envisaged for the underlying risky asset on which is defined the CPPI strategy. In particular, we investigate the CPPI strategy under rough volatility with the rough Heston model (using a power law kernel K) since this model is known for being extremely consistent with time series of historical volatility. We then show how to obtain effectively very accurate results from the characteristic function of the log-cushion, such as the moments, density, several risk measures as well as some gap risk measures for the portfolio value P_T . Moreover, the Heston model, which also belongs to the class of affine Volterra processes (with an identity kernel), is further studied in this chapter and we again get results in a more effective and accurate way using the characteristic function of the log-cushion, compared with classical simulation-based methods. For the sake of comparison, this characteristic function is also given under the simple B&S framework.

These three models are estimated directly under the real probability measure based on the PMCMC procedure of Andrieu et al. (2010) from a dataset of S&P500 daily log-returns. In particular, this PMCMC algorithm is adapted to the rough Heston case so as to take into account the non-Markovianity of its variance process. This allows to deviate from the existing calibration procedures for the rough Heston model, exclusively based on risk-neutral data, and to provide a better statistical fit of the past time series of log-returns. From this econometric estimation of the considered models, we emphasize the importance of taking into account the roughness

of volatility for providing a better modeling of the left tail of the CPPI distribution. This roughness property of volatility has indeed a strong impact on the kurtosis of the CPPI distribution and since the CPPI strategy is very sensitive to tail events, we show that the rough Heston model enables to study more accurately the true properties of the portfolio value at maturity. These effects are confirmed with the analysis of the log-cushion density under each model as well as with the study of several risk measures such as the VaR and TVaR. Finally, in a more practical way, we show that the rough Heston best models the gap risk associated with the CPPI strategy. It therefore provides a better protection of the guaranteed amount at maturity for the client and allows the issuer to build a more accurate hedging strategy.

2.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 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. If we take the power-law kernel $K(t) = t^{r-1}/\Gamma(r)$, the fractional integral and derivative are given by the following convolution operator

$$I^r f(t) = (K \star f)(t), \quad D^r f(t) = \frac{d}{dt} (I^{1-r} f(t)).$$

2.B Proof of Theorem 2.1

The proof closely follows the one of Gatheral and Keller-Ressel (2019) and expands their results in the context of CPPI. Let

$$G_s = (g \star \xi)_s(T, u) = \int_s^T g(T-r, u) \xi_s(r) dr, \quad (2.B.1)$$

$$M_s = \exp(uY_s + G_s). \quad (2.B.2)$$

Hence, it directly comes from (2.3.5) that M is a martingale. It is also shown in Gatheral and Keller-Ressel (2019) that G can be rewritten as an Itô process of the form

$$G_s = \int_0^T g(T-r, u) \xi_0(r) dr - \int_0^s g(T-r, u) V_r dr + \int_0^s h_r(T, u) dW_r, \quad (2.B.3)$$

where

$$h_s(T, u) = (g \star \eta)_s(T, u) = \int_s^T g(T-r, u) \eta_s(r) dr.$$

This equation 2.B.3 for G is obtained using the stochastic Fubini theorem which requires the *Assumption 2.1* of Gatheral and Keller-Ressel (2019) imposing the two following conditions on $\eta_s(\cdot)$:

- (a) For $dt \otimes d\mathbb{P}$ -almost all (t, ω) , it hold that $t \mapsto \eta_s(s+t, \omega)$ is non-negative and decreasing on $(0, T)$.

(b) For any $T > 0$, the following integrability condition holds almost surely :

$$\int_0^T \left(\int_0^T \eta_s(r)^2 ds \right)^{1/2} dr < \infty.$$

In particular, these conditions are satisfied in the rough Heston model with $\eta_s(t) = \sqrt{V_s} \kappa(t-s)$, as defined in (2.2.4). We now expand their results and show that the CGF is still affine when considering the following dynamic for Y (see equation (2.3.4)) :

$$dY_s = -\frac{1}{2}m^2 V_s ds + m\sqrt{V_s} dW_s.$$

Then, applying Itô's lemma to $M_s = \exp(uY_s + G_s)$, we have that

$$\begin{aligned} \frac{dM_s}{M_s} &= (dW_s \text{ terms}) \\ &+ \left(\frac{1}{2}m^2(u^2 - u)V_s - g(T-s, u)V_s + m\mu\rho\sqrt{V_s}h_s(T, u) + \frac{1}{2}h_s(T, u)^2 \right) ds. \end{aligned}$$

Since we know from Section 2.2 that $\eta_s(T) = \sqrt{V_s} \kappa(T-s)$, we find the following simplification for $h_s(T, u)$:

$$h_s(T, u) = (g \star \eta)_s(T, u) = \sqrt{V_s} \int_0^{T-s} g(T-s-r, u) \kappa(r) dr = \sqrt{V_s} (g \star \kappa)(T-s).$$

Therefore, if we define

$$R_v(m, u, w) = \frac{1}{2}m^2(u^2 - u) + m\mu uw + \frac{1}{2}w^2, \quad (2.B.4)$$

we find that

$$\frac{dM_s}{M_s} = (dW_s \text{ terms}) + (R_v(m, u, (g \star \kappa)(T-s)) - g(T-s, u)) V_s ds.$$

In order to have that M is a local martingale, we need to cancel the terms in ds and we hence find the following expression for $g(\cdot, u)$,

$$g(t, u) = R_v(m, u, (g \star \kappa)(t, u)) = R_v\left(m, u, \int_0^t \kappa(t-r)g(r, u)dr\right), \quad (2.B.5)$$

Finally, following Gatheral and Keller-Ressel (2019), one can show that M is a true martingale and then, expression 2.B.5 enables to compute the CGF of the forward variance model (Y, ξ) using

$$\log \mathbb{E} \left[e^{u(Y_T - Y_s)} | \mathcal{F}_s \right] = \int_s^T g(T-r, u) \xi_s(r) dr. \quad (2.B.6)$$

□

2.C Algorithms

Algorithm 1 Particle filtering of the rough Heston model.

Initial step : draw N values of $v_0^{(i)}$ for $i = 1, \dots, N$ from the initial distribution $f(v_0)$, which is assumed to be uniform $\mathcal{U}_{[0,0.36]}$.

Main procedure :

for $j = 1$ to n **do**

Prediction step :

 Draw a sample $\Delta w_j^{2(i)}$ from a $\mathcal{N}(0, \sqrt{\Delta})$, $i = 1, \dots, N$.

 Update $v_j^{(i)}$ using the relation :

$$v_j^{(i)} = v_0^{(i)} + \sum_{k=0}^{j-1} \frac{(t_j - t_k)^{\alpha-1}}{\Gamma(\alpha)} \lambda \left(\theta - v_k^{(i)} \right) \Delta + \sum_{k=0}^{j-1} \frac{(t_j - t_k)^{\alpha-1}}{\Gamma(\alpha)} \nu \sqrt{v_k^{(i)}} \Delta w_{k+1}^{2(i)}$$

Correction step :

 The particle $p_j^{(i)} = \left(v_k^{(i)}, \Delta w_k^{2(i)} \right)_{k=1:j}$ has a probability of occurrence

$$\beta_j^{(i)} = \frac{f(r_j | p_j^{(i)})}{\sum_{i=1:N} f(r_j | p_j^{(i)})} \quad (2.C.1)$$

where

$$f(r_j | p_j^{(i)}) = \phi \left(\frac{r_j - \left(\mu - \frac{v_{j-1}^{(i)}}{2} \right) \Delta - \rho \sqrt{v_{j-1}^{(i)}} \Delta w_j^{2(i)}}{\sqrt{1 - \rho^2} \sqrt{v_{j-1}^{(i)}} \sqrt{\Delta}} \right)$$

Resampling step :

 Resample with replacement N particles with importance weights $\beta_j^{(i)}$.

 The initial importance weights are then set to $\beta_j^{(i)} = 1/N$.

end for

Algorithm 2 Metropolis Hastings algorithm.

Main procedure :
for $k = 1$ to K **do**

 Simulate $\vartheta' \sim q(\vartheta' | \vartheta_k)$

Take

$$\vartheta_{k+1} = \begin{cases} \vartheta' & \text{with probability } \rho(\vartheta_k, \vartheta') \\ \vartheta_k & \text{with probability } 1 - \rho(\vartheta_k, \vartheta') \end{cases}$$

 where $\rho(\vartheta_k, \vartheta')$ is the acceptance probability

$$\rho(\vartheta_k, \vartheta') = \min \left\{ \frac{f(\vartheta' | r)}{f(\vartheta_k | r)} \frac{q(\vartheta_k | \vartheta')}{q(\vartheta' | \vartheta_k)}, 1 \right\}$$

end for

2.D Proof of Theorem 2.2

For an investment horizon T , we define $M_t = e^{U_t}$, with U_t given by

$$\begin{aligned} U_t &= \phi(T-t) + \chi(T)V_0 - \int_0^t \chi'(T-s)V_s ds + \int_0^t \psi_2(T-s)dL_s + \psi_1(T-s)Y_t \\ &= \phi(T-t) + \chi(T)V_0 - \int_0^t \chi'(T-s)V_s ds + \int_0^t \psi_2(T-s)dL_s + u_1 Y_t, \quad (2.D.1) \end{aligned}$$

since $\psi_1 = u_1$ due to the form (2.5.14) and where $dL_t = (\beta + bV_s) ds + (0 + a\sqrt{V_s}) d\widehat{W}_s$ as in Section 2.5.2 to ensure that the process $Z = (Y, V)^\top$ is an affine Volterra process. Using Itô's lemma on $M_t = e^{U_t}$, we have

$$\begin{aligned} \frac{dM_t}{M_t} &= -\phi'(T-t) - \chi'(T-t)V_t dt + \psi_2(T-t)(\beta + bV_t)dt + \frac{1}{2}\psi_2(T-t)^2 a^2 V_t dt \\ &\quad + \frac{1}{2}m^2(u_1^2 - u_1)V_t dt + \rho\psi_2(T-t) \left(0 + a\sqrt{V_t}\right) u_1 m \sqrt{V_t} dt + (dW_t \text{ terms}), \end{aligned}$$

which gives

$$\frac{dM_t}{M_t} = (-\phi'(T-t) + \mathcal{R}_\Phi(\psi_2(T-t)))dt + (-\chi'(T-t) + \mathcal{R}_\Psi(\psi(T-t)))V_t dt + (dW_t \text{ terms}),$$

where from Theorem 2.2

$$\mathcal{R}_\Phi(\psi_2) = \beta\psi_2,$$

and

$$\begin{aligned}\mathcal{R}_\Psi(\psi) &= \frac{1}{2}m^2(\psi_1^2 - \psi_1) + (b + m\psi_1\rho a)\psi_2 + \frac{a^2}{2}\psi_2^2 \\ &= \frac{1}{2}m^2(u_1^2 - u_1) + (b + mu_1\rho a)\psi_2 + \frac{a^2}{2}\psi_2^2.\end{aligned}$$

Hence for M to be a local martingale on $[0, T]$, we find that $\phi' = \mathcal{R}_\Phi(\psi_2)$ and that $\chi' = \mathcal{R}_\Psi(\psi)$. Imposing that $\phi(0) = 0$ and $\chi(0) = u_2$ and using the definition of U_t as well as the form (2.5.15) of ψ_2 , it is straightforward to show that

$$U_T = \phi(0) + \chi(0)V_T + u_1Y_T = u_2V_T + u_1Y_T, \quad (2.D.2)$$

Since we assume that K satisfies (2.5.10) and that $u \in \mathbb{C}^2$ with $\operatorname{Re} u_1 \in [0, 1]$ and $\operatorname{Re} u_2 \leq 0$, the Riccati-Volterra equation (2.5.15) has a unique global solution ψ_2 and M is a true martingale, as shown in Theorem 7.1. of Abi Jaber et al. (2019). Hence, since M is a true martingale, we have with $u = (u_1, u_2)$ and $Z = (Y, V)^\top$ that

$$\mathbb{E}[e^{uZ_T}] = \mathbb{E}[e^{u_2V_T + u_1Y_T}] = \mathbb{E}[M_T] = M_0 = e^{\phi(T) + \chi(T)V_0 + u_1Y_0}.$$

□

2.E Proof of Corollary 2.3

We build the exact same proof as in Theorem 2.2 based on the same assumption on K except that we now impose $u_2 = 0$ such that $\chi(0) = 0$. Hence, the terminal value U_T of the process U defined by (2.D.1) is now

$$U_T = \phi(0) + \chi(0)V_T + u_1Y_T = u_1Y_T.$$

Since M is still a true martingale when $u_2 = 0$, we have that

$$\mathbb{E}[e^{uZ_T}] = \mathbb{E}[e^{u_1Y_T}] = \mathbb{E}[e^{U_T}] = \mathbb{E}[M_T] = M_0 = e^{\phi(T) + \chi(T)V_0 + u_1Y_0},$$

and the solution ψ now becomes

$$\begin{aligned}\psi_1 &= u_1, \\ \psi_2 &= K \star \mathcal{R}_\Psi(\psi).\end{aligned} \quad (2.E.1)$$

Finally, recall from (2.3.4) that the dynamic of the log-cushion is denoted X_t and is given by

$$dX_t = (r + m(\mu - r))dt + dY_t.$$

Hence, the Fourier-Laplace transform of the affine Volterra process $\tilde{Z} = (X, V)^\top$ with $u_2 = 0$ becomes

$$\mathbb{E}[e^{u\tilde{Z}_T}] = \mathbb{E}[e^{u_1X_T}] = e^{\phi(T) + \chi(T)V_0 + u_1(X_0 + (r + m(\mu - r))t)}.$$

And finally the characteristic function of the log-cushion $\Phi_0^X(T, z)$ under the general affine Volterra process $\tilde{Z}$ is obtained by replacing u_1 by iz , $z \in \mathbb{R}$. □

2.F Fast Fourier Transform

We present here the approach used to determine numerically the density of probability of the log-cushion X_T , which then allows to obtain the VaR and TVaR of the portfolio value P_T for each model. First, the characteristic function of X_T , denoted $\Phi_0^X(T, z)$, is by definition the Fourier Transform of the density $f_T(x)$ of X_T :

$$\Phi_0^X(T, z) = \mathbb{E}[e^{izX_T}] = \int_{-\infty}^{+\infty} e^{izx} f_T(x) dx.$$

Since we know the characteristic function of X_T for the B&S, Heston and rough Heston models, we can derive the density function $f_T(x)$ using the Fourier inversion theorem which states that

$$f_T(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_0^X(T, z) e^{-izx} dz = \frac{1}{\pi} \int_0^{+\infty} \Phi_0^X(T, z) e^{-izx} dz, \quad (2.F.1)$$

where the second equality results from the symmetry of the integrand. Next, we illustrate how to derive the density function $f_T(x)$ from equation (2.F.1) via the Fast Fourier Transform (FFT). FFT is an efficient algorithm for computing the following transformation of a vector $(\alpha_m, m = 1, \dots, M-1)$ into a vector $(\beta_m, m = 1, \dots, M-1)$, where M is typically a power of 2 :

$$\beta_m = \sum_{j=0}^{M-1} \exp\left(-\frac{2\pi i}{M} jm\right) \alpha_j.$$

This method is quicker than computing the discretization of the integral (2.F.1) since the FFT algorithm computes in only $O(m \log m)$ operations (rather than $O(m^2)$). The first step of the FFT consists in discretizing the integral (2.F.1). We denote Δz the step of discretization and M the number of steps. The discretization mesh is defined by

$$\{z_j\} = \{j\Delta z \mid 0 \leq j \leq M-1\}.$$

We then need to define a mesh of discretization for the density values $f_T(x)$, spaced by Δx and counting the same number M of elements as $\{z_j\}$. The mesh of the density is then

$$\{x_m\} = \{-x_{\min} + m\Delta x \mid 0 \leq m \leq M-1\},$$

where $x_{\min} = \frac{M\Delta x}{2}$ and with the following condition on the steps of discretizations $\Delta z \Delta x = \frac{2\pi}{M}$. Therefore, we can reformulate (2.F.1) into the following form

$$f_T(x_m) = \frac{1}{\pi} \int_0^{+\infty} \Phi_0^X(T, z) e^{-izx_m} dz = \frac{1}{\pi} \sum_{j=0}^{M-1} \Phi_0^X(T, z_j) e^{-iz_j x_m} \Delta z,$$

and hence obtain

$$\underbrace{\pi f_T(x_m)}_{\beta_m} = \sum_{j=0}^{M-1} e^{-i \frac{2\pi}{M} jm} \underbrace{e^{iz_j x_{\min}} \Phi_0^X(T, z_j) \Delta z}_{\alpha_j} . \quad (2.F.2)$$

This way, we can apply the FFT with $\alpha_j = e^{iz_j x_{\min}} \Phi_0^X(T, z_j) \Delta z$ in order to obtain the output vector $\beta_m = \pi f_T(x_m)$ for $m = 0, \dots, M-1$. Finally, once we estimated the density via the FFT, we can easily derive the ε -quantile of X_T denoted x_ε . Using equation (2.7.1), we find that the VaR at level ε of the portfolio value P_T at maturity is equal to

$$VaR_\varepsilon(P_T) = e^{x_\varepsilon} + F_T . \quad (2.F.3)$$

The TVaR at level ε of P_T can then be computed by discretizing (2.7.4) :

$$TVaR_\varepsilon(P_T) = \frac{1}{\varepsilon} \sum_{x_m: x_m \leq \log(VaR_\varepsilon(P_T) - F_0 e^{rT})} e^{x_m} f_T(x_m) \Delta x + F_0 e^{rT} . \quad (2.F.4)$$

Part II

Illiquidity modeling with fractional processes

CHAPTER 3

A subdiffusive stochastic volatility jump model

We start the second part of the thesis by modeling the illiquidity process associated with low-frequency observations of asset prices using subdiffusions (or *fractional diffusions*). Indeed, subdiffusions appear as good candidates for modeling the price of illiquid assets in financial markets since they are obtained from subordination, where a standard diffusion is time-changed by an inverse subordinator representing the illiquidity process of the associated asset price. Existing subdiffusive models of asset prices are then able to capture the motionless periods in the quotes of thinly-traded assets. However, they fail at reproducing simultaneously the jumps and the time-varying random volatility observed in the price of these assets. The aim of this work is hence to propose a new model of subdiffusive asset prices reproducing the main characteristics exhibited in illiquid markets. This is done by considering a stochastic volatility jump model, time-changed by an inverse subordinator. We derive the forward fractional PDEs governing the probability density function of the introduced model and we prove that it leads to an arbitrage-free and incomplete market under a suitable change of measure. By proposing a new procedure for estimating the model parameters and using a series expansion for solving numerically the obtained fractional PDE, we are able to price various European-type derivatives on illiquid assets and to depart from the common Markovian valuation setup. This way, we show that the introduced subdiffusive stochastic volatility jump model yields consistent and reliable results in illiquid markets.

Dupret, J. L., & Hainaut, D. (2023). A subdiffusive stochastic volatility jump model. *Quantitative Finance*, 23(6), 979-1002.

3.1 Introduction

Small capitalization stocks tend to exhibit motionless periods in their share price corresponding to a lack of liquidity associated with these assets. This occurs when a stock stops being exchanged by the investors during several days, weeks or months. This feature is also common in emerging markets in which the number of partici-

pants, and thus the number of transactions, is rather low. Such periods of constant values can be observed on the upper plot of Figure 3.1.1, which depicts the price of Quadpack Industries (ALQP), a small capitalization listed in the Eurolist Group C. As Brownian motions and infinite activity Lévy processes move infinitely often in any given time interval, they are not suitable for modeling these periods of illiquidity in financial markets. A similar behavior is however observed in physical systems exhibiting subdiffusion, a well-known and established phenomenon in statistical physics. The usual models of subdiffusion are developed in terms of fractional Fokker–Planck equations (FFPE) which were first derived in Metzler and Klafter (2000) from continuous-time random walks (CTRW) with heavy-tailed waiting times/trapping events. These FFPE hence provide an appropriate procedure for describing transport dynamics with trapping periods in complex systems, such as heavy particles’ movements in Metzler and Klafter (2004). Subdiffusive dynamics can also be obtained from subordination, where a standard diffusion process is time-changed by the so-called inverse subordinator, as in Magdziarz (2009b). Therefore, the observed motionless periods without trades in financial markets can be considered as the equivalent in physical systems of the trapping events during which the subdiffusive particles get immobilized. This explains why tick-by-tick CTRW and their related FFPE have been heavily used in finance and econophysics as early as the 2000s (see Mainardi et al. (2000), Scalas et al. (2000), Bouchaud et al. (2003) or Scalas (2006)) and this justifies their ability to describe the dynamic of asset prices in illiquid markets. Fundamental theoretical results on subdiffusions and their probabilistic properties are provided by Meerschaert and Scheffler (2004), Jurliewicz et al. (2012) and Meerschaert and Scheffler (2008).

From this similitude between particles transport and illiquid asset dynamic, the class of subdiffusive Lévy processes was introduced as models of asset prices in thinly-traded markets. This class contains anomalous diffusions where the CTRW limiting diffusion consists of a Lévy process time-changed by an independent inverse-stable subordinator. Several instances of these models have already been investigated in the literature. In terms of the generating FFPE, such a class has been studied in Cartea and del Castillo-Negrete (2007). The particular case where the parent Lévy process is a compound Poisson process was introduced in Cartea and Meyer-Brandis (2010); the Brownian motion case has been investigated in Magdziarz (2009a) where it is shown that the considered model is arbitrage-free but incomplete, and where the corresponding subdiffusive Black & Scholes (B&S) formula for the fair price of European options is obtained. A general treatment when the driving noise is a generic Lévy process has been recently provided in Torricelli (2020) and in Jacquier and Torricelli (2020), along with general Fourier option pricing techniques for Lévy-type subdiffusions. Moreover, the classical models of Mainardi et al. (2000) and Scalas et al. (2000) also admit a representation of the subdiffusive Lévy form. This class of models is strongly supported by the econometric analysis of Dufour and Engle (2000), and confirmed in numerous empirical studies later on.

The evidence is that trading activity is negatively correlated with price impact and with the volatility of the asset price: the fewer the trades (longer duration) the more sluggish the price innovation, the lower the price impact and the lower the volatility, while a higher trading intensity (short duration) is coupled with high price impact, high volume and high volatility.

A second stream of literature introduces subdiffusive models of asset prices based on fractional Brownian motions (fBm). Indeed, Gu et al. (2012), Guo and Yuan (2014), Wang et al. (2012) and Shokrollahi et al. (2016) propose an extension of the subdiffusive B&S model of Magdziarz (2009a) by introducing asset dynamics based on subdiffusive fBm. They managed to build a pricing equation for European calls under various assumptions. However, since fBm are neither Markov processes nor semi-martingales, it has been shown in Cheridito (2003) that fBm models admit arbitrage in frictionless markets. This hence forces these authors to consider modified dynamics inspired from fBm (cf. Shokrollahi et al. (2016) or Guo and Yuan (2014) with mixed fBm) or to impose proportional transaction costs in a discrete setting, as in Gu et al. (2012).

We then aim in this work at enlarging the above class of Lévy subdiffusions by considering a subdiffusive model that conjugates both stochastic volatility and jumps so as to reproduce the main characteristics observed for low-liquidity stocks in financial markets. Indeed, we first observe that these thinly-traded stocks systematically exhibit jumps in the sample path of their price process. This can be interpreted by the arrival of news in the market which suddenly change the anticipations of investors and make them trade the stock again. These jumps in price can clearly be seen on Figure 3.1.1 for the share of ALQP, which exhibits similar sample paths as the majority of small capitalizations (see *e.g.* stocks listed in the Eurolist Group C). Secondly, there are no existing subdiffusive models driven by a stochastic exogenous volatility process that can be fitted to the market so as to reproduce the clustering and leverage effects occurring in practice in the volatility data time series, as observed *e.g.* in the second plot of Figure 3.1.1 representing the annualized historical volatility of ALQP. Enlarging the class of Lévy subdiffusions by introducing a subdiffusive stochastic volatility model with jumps in the price process will hence enable to take these two aspects into account simultaneously in illiquid markets while also reproducing the well-known volatility clustering effect and the observed heavy tails of the return distribution. Finally, in the context of the classical B&S model, it is well known that time changing Brownian motions by a non-Markov subordinator (such as the first hitting time of a Lévy subordinator, see *e.g.* Stanislavsky (2003) or Meerschaert and Scalas (2006)) allows to introduce the empirically observed long-memory effect in stock returns, as investigated in Ding et al. (1993). The chosen subordinator in Stanislavsky (2003) is an α -stable process so that the memory effects are characterized by only one parameter (the α parameter). We therefore consider in this Chapter 3 a mean-reverting stochastic

volatility model with price jumps, time-changed by an inverse subordinator. We will show that this so-called *subdiffusive stochastic volatility jump* (SSVJ) model is indeed arbitrage-free, that it generates realistic sample paths incorporating periods of constant values and that it allows to reproduce the true dynamic of asset prices in illiquid markets.

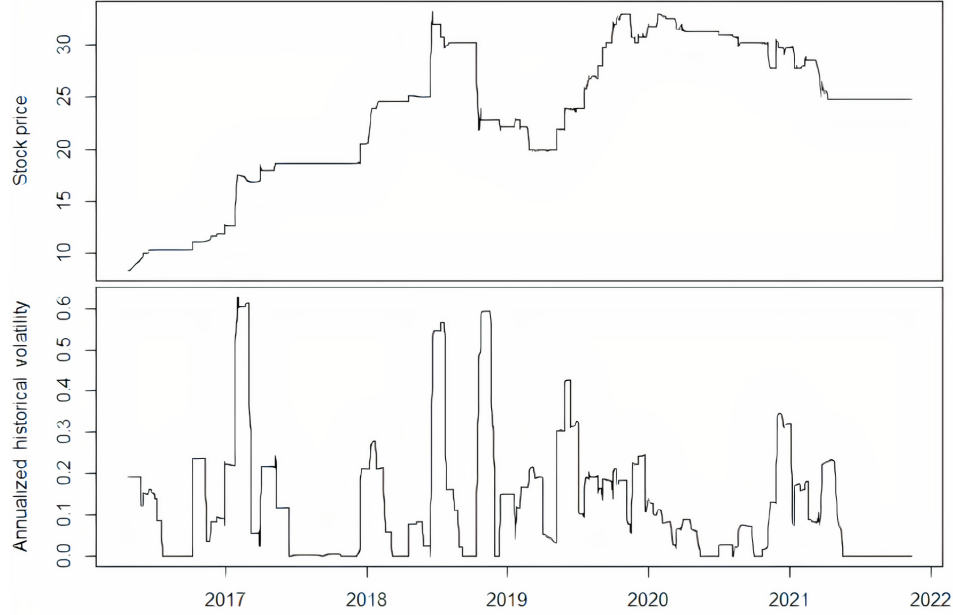


Figure 3.1.1: The upper plot depicts the stock price of ALQP over a 5-year period. The bottom plot represents the annualized historical volatility of ALQP computed using a monthly time window.

As the approach we propose in this chapter focuses on low-frequency observations of illiquid asset prices, we neglect the bid-ask spread observed in practice. This is however not the only possibility to deal with illiquidity in asset price modeling. One may think of a high-frequency trading approach, see Chapter 5. Such an approach would consist in explaining the illiquidity of assets by either the absence of buy or sell orders in the market, or by the existence of a large bid-ask spread in the limit order book, see Cartea et al. (2015) or Bacry and Muzy (2014). A possible way to introduce illiquidity in such a high-frequency approach could be to time-change the arrival of orders by an inverse subordinator. Another idea would be to directly fit this SSVJ model to high-frequency data where jumps would represent the occurrence of (buy/sell) market orders, the constant parts of the process would model the periods without trades and the diffusion part would state for the noise.

The contributions of this chapter are multiple. We first show how to price financial derivatives under the SSVJ model by deriving new FFPE that allow to describe the behavior of this newly introduced model. We generalize these forward fractional equations to a whole class of Lévy subordinators (with no drift) and we show that they are also valid if seen from a later time after inception, conditionally to some information available at that time. This is achieved by proposing a new estimation procedure that allows to depart from the common Markovian valuation setup. We next define a change of measure and from it, we show that we can obtain a risk-neutral measure $\mathbb{Q}$ for our model, leading to an arbitrage-free and incomplete market. Based on this risk-neutral measure and on the FFPE satisfied by the SSVJ model, we are able to price various European-type financial products (at inception and at any subsequent times). However, even if subdiffusions appear as good candidates for modeling illiquidity, option pricing and risk management in this framework remain a challenging task. We therefore show in the practical part of the chapter based on a α -stable subordinator how to solve numerically via a series expansion the FFPE driving the (Fourier-Laplace transform of) the bi-variate probability density of the subdiffusive log-stock price and its time-changed variance process. Finally, we contribute with this chapter to provide a full econometric estimation of the subdiffusive SVJ parameters based on the time series of observed log-returns using the Particle Monte-Carlo Markov Chain algorithm. So far, only the parameter driving the motionless periods of illiquid asset prices has been estimated in Janczura and Wyłomańska (2009) and we therefore propose a new procedure for evaluating simultaneously all of the model parameters.

3.2 A Fokker-Planck equation for jump-diffusion models

We start by deriving the classical Fokker-Planck equations (FPE) satisfied by the SVJ model in the non-subdiffusive case. From these FPE, we will manage in the sequel of this chapter to build their fractional counterpart so as to describe the behavior of the subdiffusive SVJ model. We hence consider in a first step the following classical jump-diffusion model¹ under the physical measure $\mathbb{P}$:

$$\frac{dA_t}{A_{t-}} = \mu dt + \sqrt{V_t} dW_{1,t} + (e^J - 1) dN_t, \quad (3.2.1)$$

$$dV_t = \kappa(\gamma - V_t) dt + \sigma \sqrt{V_t} dW_{2,t}, \quad (3.2.2)$$

where $A_{t-} = \lim_{s \rightarrow t} A_s$ is the left limit of the price process (so as to ensure the predictability of the integrand). Here and in the rest of the article, W_j ($j = 1, 2$), are two correlated Brownian motions defined on $(\Omega, \mathcal{F}^A, \mathbb{P}^A := (\mathcal{F}_t^A)_{t \geq 0}, \mathbb{P})$ such that $\text{Cov}[W_{1,t} W_{2,t}] = \rho t$. We define $N = (N_t)_{t \geq 0}$ as a Poisson process with

¹ A_t denotes in this chapter the asset price at time t since S_t is used for the inverse subordinator. The Heston parameters κ, γ, σ are also adapted compared with Chapters 1 and 2.

intensity parameter $\lambda_J > 0$ (i.e. $\mathbb{E}^\mathbb{P}[N_t] = \lambda_J t$) and which is independent from W_1 and W_2 . The random variable J with probability density function (pdf) $f_J(\cdot)$ and support in $\mathbb{R}$ describes the magnitude of the jump when it occurs and is assumed to be $\mathcal{F}_0^A$ -measurable. Finally, we need to impose $\mathbb{E}^\mathbb{P}[e^{2J}] < \infty$ such as to have the desired (square-integrable) martingale property of the discounted price process $(e^{-rt} A_t)_{t \geq 0}$ under the risk-neutral version of the SVJ model, see Propositions 8.8 and 9.8 of Tankov (2003). Note that all the results in the sequel of this chapter could be extended using a more general square integrable Lévy process in the jump diffusion (3.2.1). Applying Itô's Lemma for semimartingales (Proposition 8.19 of Tankov (2003)) since equation (3.2.1) is the Dolean-Dades exponential of a Lévy process, we find the following dynamic for the log-price $X_t := \ln A_t$,

$$dX_t = \left(\mu - \frac{1}{2} V_t \right) dt + \sqrt{V_t} dW_{1,t} + J dN_t. \quad (3.2.3)$$

Hence, the specification of the SVJ model implies that the jump size of the log-price process $X_t - X_{t-}$ is equal to the random variable J , when a jump occurs. Let us write $p(t, x, v | s, x_s, v_s)$ the conditional bivariate pdf of $(X_t, V_t | X_s = x_s, V_s = v_s)$. This function is such that

$$p(t, x, v | s, x_s, v_s) := \frac{\partial^2}{\partial x \partial v} \mathbb{P}(X_t \leq x, V_t \leq v | X_s = x_s, V_s = v_s). \quad (3.2.4)$$

This pdf is solution of a forward partial differential equation (PDE) as derived *e.g.* in Bates (1996) and follows directly from the standard theory of infinitesimal generator for jump diffusions (and more generally affine diffusions). More precisely, the conditional bivariate pdf $p(t, x, v | s, x_s, v_s)$ at time $t \geq s$ of the pair (X_t, V_t) is solution of the following Fokker-Planck equation (FPE) :

$$\begin{aligned} \frac{\partial}{\partial t} p(t, x, v | \cdot) = & -\frac{\partial}{\partial x} \left[\left(\mu - \frac{1}{2} v \right) p(t, x, v | \cdot) \right] - \frac{\partial}{\partial v} [\kappa (\gamma - v) p(t, x, v | \cdot)] \\ & + \frac{1}{2} \frac{\partial^2}{\partial x^2} [v p(t, x, v | \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial v^2} [\sigma^2 v p(t, x, v | \cdot)] + \frac{\partial^2}{\partial x \partial v} [\rho \sigma v p(t, x, v | \cdot)] \\ & + \lambda_J \int_{\mathbb{R}} p(t, x - j, v | \cdot) f_J(j) dj - \lambda_J p(t, x, v | \cdot) \end{aligned} \quad (3.2.5)$$

with initial condition $p(s, x, v | s, x_s, v_s) = \delta_{\{x=x_s, v=v_s\}}$, where $\delta_{\{z\}}$ is the Dirac measure located at z . This FPE is hard to solve due to the initial condition involving a Dirac measure. Numerical methods for computing such forward PDE indeed fail to be reliable since the initial condition cannot be evaluated properly. We hence focus on the forward PDE of the Fourier-Laplace transform of the density, which is way simpler than equation (3.2.5) and which will allow to obtain more stable numerical results since its initial condition does not rely on a Dirac delta anymore.

Summarizing Drăgulescu and Yakovenko (2002), let us first denote by $\phi(t, z_1, v |$

s, x_s, v_s) the Fourier transform with respect to x of the conditional bivariate pdf $p(t, x, v \mid s, x_s, v_s)$ of the pair (X_t, V_t) . This Fourier transform is defined as follows with $z_1 \in \mathbb{R}$:

$$\phi(t, z_1, v \mid s, x_s, v_s) := \int_{-\infty}^{\infty} e^{-iz_1 x} p(t, x, v \mid s, x_s, v_s) dx,$$

and we thus have that the inverse Fourier transform is given by

$$p(t, x, v \mid s, x_s, v_s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{iz_1 x} \phi(t, z_1, v \mid s, x_s, v_s) dz_1. \quad (3.2.6)$$

Inserting (3.2.6) into (3.2.5) and simplifying the known information $\{X_s = x_s, V_s = v_s\}$ by $\{\cdot\}$, we have

$$\begin{aligned} \frac{\partial}{\partial t} \phi(t, z_1, v \mid \cdot) &= -\frac{\partial}{\partial v} [\kappa(\gamma - v) \phi(t, z_1, v \mid \cdot)] + \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \phi(t, z_1, v \mid \cdot) \\ &\quad - \left[\frac{z_1^2}{2} v + iz_1 \left(\mu - \frac{1}{2} v \right) - i\rho\sigma z_1 \frac{\partial}{\partial v} v - \frac{\sigma^2}{2} \frac{\partial^2}{\partial v^2} v \right] \phi(t, z_1, v \mid \cdot), \end{aligned} \quad (3.2.7)$$

where $\hat{f}_j(z_1)$ is the Fourier transform of $f_j(\cdot)$. The equation (3.2.7) is simpler than (3.2.5) because the number of variables has been reduced to two, v and t , whereas z_1 only plays the role of parameter. Finally, since v is positive and since (3.2.7) is linear in v and quadratic in $\frac{\partial}{\partial v}$, it can then be simplified by taking the Laplace transform over v :

$$\phi(t, z_1, z_2 \mid s, x_s, v_s) = \int_0^{+\infty} e^{-z_2 v} \phi(t, z_1, v \mid s, x_s, v_s) dv. \quad (3.2.8)$$

We therefore retrieve the following well-known result in the theory of jump diffusions. The Fourier-Laplace transform $\phi(t, z_1, z_2 \mid \cdot)$ of the joint pdf $p(t, x, v \mid \cdot)$ in the SVJ model is solution of the following classical forward PDE

$$\begin{aligned} \frac{\partial}{\partial t} \phi(t, z_1, z_2 \mid \cdot) &= - \left(\kappa\gamma z_2 + iz_1\mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi(t, z_1, z_2 \mid \cdot) \\ &\quad + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - iz_1}{2} \right) \frac{\partial}{\partial z_2} \phi(t, z_1, z_2 \mid \cdot), \end{aligned} \quad (3.2.9)$$

with initial condition $\phi(s, z_1, z_2 \mid s, x_s, v_s) = e^{-iz_1 x_s - z_2 v_s}$ and boundary condition $\phi(t, 0, 0 \mid s, x_s, v_s) = 1$. This first-order partial differential equation is solved analytically in Drăgulescu and Yakovenko (2002). Their analytical expression is however not useful in the subdiffusive case, as we will see in Section 3.4.

3.3 Subordinators

Subordinators are $\mathbb{R}^+$ -valued stochastic processes, denoted $U = (U_t)_{t \geq 0}$, with non-decreasing sample paths and starting from $U_0 = 0$. Let U be defined on the probability space $(\Omega, \mathcal{F}^U, \mathbb{P})$. One of their most important probabilistic applications

is their use as stochastic clock for time-changed processes. Moreover, we consider Lévy subordinators for which increments are independent and homogeneously distributed. Interested readers may refer to Applebaum (2009) for an introduction to such Lévy subordinators. We then have from the Lévy-Khintchine formula that the Laplace transform of Lévy subordinators has the following form

$$\mathbb{E}^{\mathbb{P}}(e^{-\lambda U_t}) = e^{-tf(\lambda)}, \quad (3.3.1)$$

where $f(\lambda) = b\lambda + \int_0^\infty (1 - e^{-\lambda z}) \nu(dz)$ and $b \geq 0$. The function $\nu(\cdot)$ is a non-negative measure on $(0, \infty)$, known as the Lévy measure and satisfying the requirement $\int_0^\infty (z \wedge 1) \nu(dz) < \infty$. The inverse of a subordinator is a process denoted by $S = (S_t)_{t \geq 0}$ and defined by

$$S_t := \inf \{ \tau \geq 0 : U_\tau > t \}. \quad (3.3.2)$$

This is the first passage time of U above the level t . This inverted Lévy subordinator is in general no longer a Lévy process. However, S is still positive and non-decreasing and hence, it has all the required properties to be used as a stochastic clock. Moreover, S is an almost surely (a.s.) locally bounded family of stopping times, diverging a.s. to infinity for large times, which is desirable in order to be a realistic random clock (and more precisely a time change, as defined in Chapter 10 of Jacod (1979)). We only consider in the sequel of this chapter a.s. strictly increasing subordinators U with infinite activity $\int_0^\infty \nu(dz) = \infty$ since in this case, the inverse subordinator S is a.s. continuous². Moreover, by construction, the inverted process S may be constant over some intervals of time. Therefore, any process subordinated by S tends to exhibit motionless periods. The probability density functions of U_t and S_t are respectively denoted by $p_U(\tau, t) = \frac{\partial}{\partial t} \mathbb{P}(U_\tau \leq t)$ and $g(t, \tau) = \frac{\partial}{\partial \tau} \mathbb{P}(S_t \leq \tau)$. By definition, we have that $\mathbb{P}(S_t > s) = \mathbb{P}(U_s < t)$. In order to use our SSVJ model for risk management and hedging as well as to price financial derivatives after their inception, we need to establish the features of $S_t - S_s$ conditionally to the available information up to time s . Most of the literature focuses on the properties of S seen from time 0, when $U_0 = S_0 = 0$. In an actuarial context with α -stable subordinators, Hainaut (2021) extends these properties when the process S is considered from a later time s . However, S_t is by nature non-Markovian and its natural filtration up to time $s < t$ is not sufficient to draw any conclusion about $S_t - S_s$. We hence need to adjunct additional information about the subordinator that is inverted. It is shown in this paper that most of the properties available at time 0 admit an extension at later time s if the available information is the value of S_s and U_{S_s} , where U_S is known in Bertoin (1999) as the first-passage process. We can now easily extend the results of Hainaut (2021) when considering general subordinators of the form (3.3.1). First, let us denote the natural filtration of S by $\mathbb{F}^S = (\mathcal{F}_t^S)_{t \geq 0}$, the natural filtration of U by $\mathbb{F}^U = (\mathcal{F}_t^U)_{t \geq 0}$

²This assumption can be relaxed by considering lazy clocks as subordinator U since it keeps the martingale (and hence semimartingale) property of the time-changed process needed in the following sections, see Profeta and Vrins (2019). This should be studied in further research.

and the natural filtration of (S, U_S) by $\mathbb{F}^{S, U \circ S} = (\mathcal{F}_t^{S, U \circ S})_{t \geq 0}$, which are assumed to be independent from $\mathbb{F}^A$. It is clear from the definition (3.3.2) and from Jacod (1979) that S_t is a stopping-time for any $t \geq 0$ with respect to the filtration $\mathbb{F}^U$ and we can hence introduce the following stopped sigma-algebra

$$\mathcal{F}_{S_t}^U := \{A \in \mathcal{F}_\infty^U : A \cap \{S_t \leq u\} \in \mathcal{F}_u^U \text{ for all } u \geq 0\},$$

where $\mathcal{F}_\infty^U = \bigvee_{t \geq 0} \mathcal{F}_t^U$. We then have the following proposition.

Proposition 3.1. (i) For all $t \geq 0$, $\mathcal{F}_t^S \subset \mathcal{F}_{S_t}^U$ and $\mathcal{F}_t^{S, U \circ S} = \mathcal{F}_{S_t}^U$.

(ii) The bivariate process (S, U_S) is a Markov process with respect to its natural filtration $\mathbb{F}^{S, U \circ S}$.

(iii) Let $t > 0$ and $x_1 \geq 0$. Then,

$$\{U_{S_t} \geq x_1\} = \begin{cases} \Omega & \text{if } x_1 \leq t, \\ \{S_t = S_{x_1}\} & \text{if } x_1 > t. \end{cases}$$

Proof. (i) and (ii) are standard and can be found in e.g. Meerschaert and Straka (2014) or Scalas and Toaldo (2021). The proof of (iii) is relegated in Appendix 3.D. $\square$

The implications of this proposition are of utmost importance. The Markov property of (ii) is key to obtain all the following results of this section by conditioning only on the last values of S_s, U_{S_s} at time $s < t$, instead on the whole sigma-algebra $\mathcal{F}_s^{S, U \circ S}$, i.e. :

$$\mathbb{P}(S_t \in A \mid \mathcal{F}_s^{S, U \circ S}) = \mathbb{P}(S_t \in A \mid S_s, U_{S_s}),$$

for $A \in \mathcal{B}_{\mathbb{R}^+}$, $t > s$. However, property (i) implies that (S, U_S) is not $\mathbb{F}^S$ -adapted and that we are conditioning on values that are unobserved at time s . Indeed, from (iii) we have that $U_{S_s} \geq s$. Moreover, with $t > s$, we either have $U_{S_s} \geq t$, leading trivially from (iii) to the relationship $S_s = S_t$, or have $U_{S_s} < t$, leading to an absolutely continuous framework as we will see with the main result of Section 3.4. Hence, we see that at time s , it is unknown whether it will be $U_{S_s} < t$ or $U_{S_s} \geq t$ and whether we will work in the absolutely continuous framework or in the frozen one. This issue is addressed in Section 3.6 through the estimation procedure which allows to retrieve the pair (S_s, U_{S_s}) from the observed illiquid data up to time s .

As just mentioned, if $U_{S_s} = u \geq t > s \geq 0$, the clock is stopped between s and t and we get trivially $\mathbb{P}(S_t - S_s \leq \tau \mid S_s = w, U_{S_s} = u) = 1$ for all $\tau \geq 0$. We therefore consider in the sequel of this section only the non-frozen case $U_{S_s} = u < t$. We have for $0 \leq s \leq u < t$ that the cdf of $S_t - S_s$ is related to the cdf of U_t through the relation

$$\begin{aligned} \mathbb{P}(S_t - S_s \leq \tau \mid S_s = w, U_{S_s} = u) &= \mathbb{P}(U_{\tau+w} \geq t \mid S_s = w, U_{S_s} = u) \\ &= \mathbb{P}(U_\tau \geq t - u). \end{aligned}$$

We next find that the conditional pdf of $S_t - S_s$, denoted by $g(t, \tau \mid s, w, u)$ for $t > u \geq s \geq 0$, is the derivative of this last expression with respect to τ :

$$\begin{aligned} g(t, \tau \mid s, w, u) &= \frac{\partial}{\partial \tau} \mathbb{P}(S_t - S_s \leq \tau \mid S_s = w, U_{S_s} = u) = -\frac{\partial}{\partial \tau} \mathbb{P}(U_\tau \leq t - u) \\ &= -\frac{\partial}{\partial \tau} \int_0^{t-u} p_U(\tau, x) dx. \end{aligned}$$

The Laplace transform of $g(t, \tau \mid s, w, u)$ with respect to time t , denoted $\tilde{g}(\lambda, \tau \mid s, w, u)$, is therefore equal to

$$\tilde{g}(\lambda, \tau \mid s, w, u) = -\frac{\partial}{\partial \tau} \int_u^\infty e^{-\lambda t} \int_0^{t-u} p_U(\tau, x) dx dt. \quad (3.3.3)$$

The subordinator U being strictly positive, its density is null on $\mathbb{R}^-$ and integrating by parts gives

$$\tilde{g}(\lambda, \tau \mid s, w, u) = -\frac{\partial}{\partial \tau} \left(\frac{1}{\lambda} e^{-\lambda u} \tilde{p}_U(\tau, \lambda) \right) = \frac{f(\lambda)}{\lambda} e^{-\lambda u - \tau f(\lambda)}, \quad (3.3.4)$$

since from equation (3.3.1), we have $\tilde{p}_U(\tau, \lambda) = \mathbb{E} \mathbb{P}(e^{-\lambda U_\tau}) = e^{-\tau f(\lambda)}$.

We now introduce the Dzerbayshan-Caputo (D-C) derivative. First, a Bernstein function is a function $f : (0, \infty) \rightarrow \mathbb{R}$ of class C^∞ such that $f(x) \geq 0$ for all $x > 0$ and for which $(-1)^k f^{(k)}(x) \leq 0$, $\forall x > 0$, $k \in \mathbb{N}$. A Bernstein function admits a similar representation as the Laplace exponent of a Lévy subordinator with $U_0 = a$,

$$f(x) = a + bx + \int_0^\infty (1 - e^{-xz}) \nu(dz), \quad (3.3.5)$$

where $a, b \geq 0$ (see Schilling et al. (2010) for the other representations of Bernstein functions). As above, ν is a non-negative Lévy measure on $(0, \infty)$. The triplet (a, b, ν) is the Lévy triplet of the Bernstein function. We then denote by $\bar{\nu}(s)ds$ the Lévy tail measure defined in Chapter 1 of Bertoin (1999) by

$$\bar{\nu}(s)ds := \left(a + \int_s^\infty \nu(dz) \right) ds. \quad (3.3.6)$$

Observe from Section 2 of Toaldo (2015) that $\bar{\nu}$ is a right-continuous and non-increasing function for which

$$\int_0^1 \bar{\nu}(s)ds < \infty \quad \text{and} \quad \int_s^\infty \nu(dz) < \infty, \quad \forall s > 0. \quad (3.3.7)$$

Let us consider f a Bernstein function, and its Lévy tail measure $\bar{\nu}(s)ds$ that is absolutely continuous on $(0, \infty)$. Let also a function $h \in AC([u, +\infty))$, that is the

set of absolutely continuous function on $[u, +\infty)$. The generalized D-C derivative of h according to the Bernstein function f is then defined as in Toaldo (2015) by

$${}^f\mathcal{D}_t^{(u, +\infty)} h(t) := b \frac{d}{dt} h(t) + \int_0^{t-u} \frac{\partial}{\partial t} h(t-s) \bar{\nu}(s) ds, \quad t \in [u, +\infty). \quad (3.3.8)$$

This integral is well defined if $|h(t)| \leq M e^{\lambda_0 t}$ for some $\lambda_0, M > 0$. From now on, we will only consider subordinators with Lévy triplet $(0, 0, \nu)$ associated with Bernstein functions of the form $f(x) = \int_0^\infty (1 - e^{-xz}) \nu(dz)$ and Lévy tail measures $\bar{\nu}(s) ds = \int_s^\infty \nu(dz) ds$. This results from the following lemma, which is an extension of Lemma 2.5 of Toaldo (2015), where the generalized D-C derivative is here defined instead on $[u, +\infty)$ with $u \geq 0$, implying the condition $a = b = 0$.

Lemma 3.1. *Let ${}^f\mathcal{D}_t^{(u, +\infty)} h(t)$, $t > u \geq 0$ be defined as in equation (3.3.8), where f is a Bernstein function with Lévy triplet $(0, 0, \nu)$ and a Lévy tail measure $\bar{\nu}(s) ds = \int_s^\infty \nu(dz) ds$. Then the Laplace transform $\mathcal{L}$ of the generalized D-C derivative of h according to f is given by*

$$\mathcal{L} \left[{}^f\mathcal{D}_t^{(u, +\infty)} h(t) \right] (\lambda) = f(\lambda) \tilde{h}(\lambda) - \frac{f(\lambda)}{\lambda} e^{-\lambda u} h(u), \quad \Re(\lambda) > \lambda_0. \quad (3.3.9)$$

Proof. Taking $b = 0$ in the D-C derivative (3.3.8) with $t > u$, we find using Fubini's theorem

$$\begin{aligned} \mathcal{L} \left[{}^f\mathcal{D}_t^{(u, +\infty)} h(t) \right] (\lambda) &= \int_u^\infty e^{-\lambda t} \int_0^{t-u} \frac{\partial}{\partial t} h(t-s) \bar{\nu}(s) ds dt \\ &= \int_u^\infty e^{-\lambda t} \int_u^t \frac{\partial}{\partial z} h(z) \bar{\nu}(t-z) dz dt \\ &= \int_u^\infty \frac{\partial}{\partial z} h(z) \int_z^\infty e^{-\lambda t} \bar{\nu}(t-z) dt dz \\ &= \int_u^\infty \frac{\partial}{\partial z} h(z) e^{-\lambda z} \int_0^\infty e^{-\lambda t'} \bar{\nu}(t') dt' dz \\ &= \int_u^\infty \frac{\partial}{\partial z} h(z) e^{-\lambda z} \tilde{\nu}(\lambda) dz \end{aligned}$$

where $\tilde{\nu}(\lambda)$ is the Laplace transform of the Lévy tail measure (3.3.6). Fubini's theorem is justified by the absolute continuity of h and the relations (3.3.7). Integrating by parts, we find

$$\begin{aligned} \mathcal{L} \left[{}^f\mathcal{D}_t^{(u, +\infty)} h(t) \right] (\lambda) &= [h(z) \tilde{\nu}(\lambda) e^{-\lambda z}]_{z=u}^\infty + \tilde{\nu}(\lambda) \lambda \int_u^\infty e^{-\lambda z} h(z) dz \\ &= \tilde{\nu}(\lambda) \lambda \tilde{h}(\lambda) - h(u) \tilde{\nu}(\lambda) e^{-\lambda u} \end{aligned}$$

Now, taking the Laplace transform of the Lévy tail measure (3.3.6) with $a = 0$, we have

$$\tilde{\nu}(\lambda) = \int_0^\infty e^{-\lambda s} \int_s^\infty \nu(dz) ds = \frac{1}{\lambda} \int_0^\infty (1 - e^{\lambda s}) \nu(ds) = \frac{f(\lambda)}{\lambda}, \quad (3.3.10)$$

where $f(\lambda)$ is the Bernstein function given in (3.3.5) with Lévy triplet $(0, 0, \nu)$, see Section 1.2 of Bertoin (1999). Hence, we find $\tilde{\nu}(\lambda) = f(\lambda)/\lambda$, which finally leads to (3.3.9). $\square$

We will therefore only study in the sequel of the chapter strictly increasing pure-jump subordinators starting at $U_0 = 0$ with no drift and infinite activity. Nevertheless, this class of subordinators encompasses most of the subordinators of interest for modeling financial illiquidity such as α -stable and Gamma subordinators.

We now extend the forward PDE (3.2.5)-(3.2.9) satisfied by the probability density $p(t, x, v | \cdot)$ and its Fourier-Laplace transform $\phi(t, z_1, z_2 | \cdot)$ in a subdiffusive context. We show in this subdiffusive setting that they also obey a forward PDE but with respect to a D-C derivative instead of a classical derivative.

3.4 A Fokker-Planck equation for the subdiffusive SVJ (SSVJ) model

In the fractional setting of the SSVJ model, the dynamic of the subdiffusive asset price A_S (time-changed by an independent inverse subordinator S) is given under the physical measure $\mathbb{P}$ by

$$\frac{dA_{S_t}}{A_{S_{t-}}} = \mu dS_t + \sqrt{V_{S_t}} dW_{1,S_t} + (e^J - 1) dN_{S_t}, \quad (3.4.1)$$

$$dV_{S_t} = \kappa(\gamma - V_{S_t}) dS_t + \sigma \sqrt{V_{S_t}} dW_{2,S_t}. \quad (3.4.2)$$

with left limit $A_{S_{t-}} = \lim_{s \rightarrow t} A_{S_s}$. We will explain in Section 3.5 how an equivalent risk-neutral measure $\mathbb{Q}$ can be built from $\mathbb{P}$ using a time-changed Radon-Nikodym derivative. We first recall that the natural filtration of the asset price process $A = (A_t)_{t \geq 0}$ in equation (3.2.1) is denoted $\mathbb{F}^A$ and is independent from the natural filtration $\mathbb{F}^S$ of the inverse Lévy subordinator S . We then define the augmented filtration $\mathbb{F}^{A,U} := \mathbb{F}^A \vee \mathbb{F}^U$. Since S_t is a $\mathbb{F}^U$ -stopping time for all $t \geq 0$, it is also a $\mathbb{F}^{A,U}$ -stopping time. Therefore, we can introduce the filtration $\mathbb{F}^{A \circ S}$ defined as the time-changed filtration $\mathbb{F}^{A \circ S} := (\mathcal{F}_{S_t}^{A,U})_{t \geq 0}$, which coincides with the natural filtration $\mathbb{F}^{A,S,U \circ S} = \mathbb{F}^A \vee \mathbb{F}^{S,U \circ S}$, using (i) of Proposition 3.1. Since the process A is a $\mathbb{F}^{A,U}$ -semimartingale (here a Doleans-Dades exponential) and since $S = (S_t)_{t \geq 0}$ is an a.s. finite and continuous $\mathbb{F}^{A,U}$ -time change, then from Corollary 10.12 of Jacod (1979), the time-changed process $A_S = (A_{S_t})_{t \geq 0}$ is also a $\mathbb{F}^{A \circ S}$ -semimartingale,

which justifies the use of the Itô Lemma for semimartingales in the sequel³. Hence, we see that the filtration of the asset price A_S can be chosen to be $\mathbb{F}^{A \circ S}$. Finally, from Proposition 3.1 (ii) and from Meerschaert and Straka (2014), we have directly that (A_S, U_S) is a Markov process with respect to $\mathbb{F}^{A \circ S}$, which justifies in what follows the conditioning only on the values (A_{S_s}, U_{S_s}) at time s and not on the whole sigma-algebra $\mathcal{F}_s^{A \circ S}$.

We also emphasize that it appears logical to subordinate both the price process and the variance process by the same inverse subordinator S_t since when the price process is frozen, the variance process does not matter anymore and can be considered as being fixed during these periods of illiquidity. As in the classical SVJ model, taking the time-changed log-price $X_{S_t} := \ln A_{S_t}$ instead of A_{S_t} will allow to drastically simplify the FFPE we will have to solve. In the subdiffusive case, we obtain the following dynamic for the log-price via the Itô's Lemma for semimartingales applied to $\ln(A_{S_t})$:

$$dX_{S_t} = \left(\mu - \frac{1}{2} V_{S_t} \right) dS_t + \sqrt{V_{S_t}} dW_{1, S_t} + J dN_{S_t}. \quad (3.4.3)$$

The bivariate pdf of (X_{S_t}, V_{S_t}) , conditionally to $\{S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s\}$, is denoted by $p_S(t, x, v \mid s, w, u, x_s, v_s)$ and defined for $t > s$ by

$$\begin{aligned} p_S(t, x, v \mid s, w, u, x_s, v_s) \\ := \frac{\partial^2}{\partial x \partial v} \mathbb{P}(X_{S_t} \leq x, V_{S_t} \leq v \mid S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s), \end{aligned}$$

This bivariate pdf is the solution of a forward fractional PDE as stated in the next proposition.

Proposition 3.2. *For $0 \leq s \leq u < t$, the bivariate pdf p_S of (X_{S_t}, V_{S_t}) , conditionally to $\{S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s\}$, is solution of the following fractional Fokker-Planck equation (FFPE) :*

$$\begin{aligned} {}^f\mathcal{D}_t^{(u, +\infty)} p_S(t, x, v \mid \cdot) &= -\frac{\partial}{\partial x} \left[\left(\mu - \frac{1}{2} v \right) p_S(t, x, v \mid \cdot) \right] - \frac{\partial}{\partial v} [\kappa(\gamma - v) p_S(t, x, v \mid \cdot)] \\ &+ \frac{1}{2} \frac{\partial^2}{\partial x^2} [v p_S(t, x, v \mid \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial v^2} [\sigma^2 v p_S(t, x, v \mid \cdot)] + \frac{\partial^2}{\partial x \partial v} [\rho \sigma v p_S(t, x, v \mid \cdot)] \\ &+ \lambda_J \int_{\mathbb{R}} p_S(t, x - j, v \mid \cdot) f_J(j) dj - \lambda_J p_S(t, x, v \mid \cdot) \end{aligned} \quad (3.4.4)$$

with the initial condition $p_S(u, x, v \mid s, w, u, x_s, v_s) = \delta_{\{x=x_s, v=v_s\}}$. For all times t such that $0 \leq s < t \leq u$, then $p_S(t, x, v \mid s, w, u, x_s, v_s) = \delta_{\{x=x_s, v=v_s\}}$.

³Similar argument can be used for V_{S_t} . We refer to Kobayashi (2011) for a more thorough discussion on stochastic calculus for integrals driven by a time-changed semimartingale (since S is a.s. continuous in our context, any process is synchronized with S so that Lemma 2.3 of the above paper applies).

Proof. When $t \leq u$, we have from Proposition 3.1 (iii) that $S_t = S_s$ and hence $p_S(t, x, v \mid s, w, u, x_s, v_s) = \delta_{\{x=x_s, v=v_s\}}$. We now treat the case $t > u$ in the sequel of this proof. To lighten developments, we momentarily adopt the notations $p_S(t, x, v \mid \cdot) := p_S(t, x, v \mid s, w, u, x_s, v_s)$ and $g(t, \tau \mid \cdot) = g(t, \tau \mid s, w, u)$. As $g(t, \tau \mid \cdot)$ is the pdf of $S_t - S_s \mid S_s, U_{S_s}$ and from the independence between (X_t, V_t) and S_t , we infer for $0 \leq s \leq u < t$ that

$$p_S(t, x, v \mid \cdot) = \int_0^\infty p(w + \tau, x, v \mid w, x_s, v_s) g(t, \tau \mid \cdot) d\tau. \quad (3.4.5)$$

Notice that since the SVJ model (3.2.1)-(3.2.2) is Markovian,

$$p(w + \tau, x, v \mid w, x_s, v_s) = p(\tau, x, v \mid 0, x_s, v_s). \quad (3.4.6)$$

Using Fubini's theorem (the integrand in (3.4.8) being non-negative and the Lebesgue measure σ -finite), the Laplace transform of $p_S(t, x, v \mid \cdot)$ with respect to time $t > u$ is thus given by

$$\tilde{p}_S(\lambda, x, v \mid \cdot) = \int_u^\infty e^{-\lambda t} p_S(t, x, v \mid \cdot) dt \quad (3.4.7)$$

$$\begin{aligned} &= \int_u^\infty \int_0^\infty p(\tau, x, v \mid 0, x_s, v_s) e^{-\lambda t} g(t, \tau \mid \cdot) d\tau dt \\ &= \int_0^\infty p(\tau, x, v \mid 0, x_s, v_s) \tilde{g}(\lambda, \tau \mid \cdot) d\tau, \end{aligned} \quad (3.4.8)$$

where $\tilde{g}(\lambda, \tau \mid \cdot) = \int_u^\infty e^{-\lambda t} g(t, \tau \mid \cdot) dt$ is the Laplace transform of $g(t, \tau \mid \cdot)$ with respect to time t . Since this transform is given by equation (3.3.4), $\tilde{p}_S(\lambda, x, v \mid \cdot)$ is equal to

$$\begin{aligned} \tilde{p}_S(\lambda, x, v \mid \cdot) &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} \int_0^\infty p(\tau, x, v \mid 0, x_s, v_s) e^{-\tau f(\lambda)} d\tau \\ &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} \tilde{p}(f(\lambda), x, v \mid \cdot), \end{aligned} \quad (3.4.9)$$

where $\tilde{p}(\lambda, x, v \mid \cdot)$ denotes the Laplace transform of $p(t, x, v \mid 0, x_s, v_s)$ with respect to time t and with Laplace exponent λ . Let us remark that (3.4.9) is true for any Feller process X with transition law p . Applying the Laplace transform on both side of the PDE (3.2.5), we deduce that $\tilde{p}(\lambda, x, v \mid \cdot)$ is also solution of

$$\begin{aligned} \lambda \tilde{p}(\lambda, x, v \mid \cdot) - p(0, x, v \mid 0, x_s, v_s) &= -\frac{\partial}{\partial x} \left[\left(\mu - \frac{1}{2}v \right) \tilde{p}(\lambda, x, v \mid \cdot) \right] \\ &\quad - \frac{\partial}{\partial v} [\kappa(\gamma - v) \tilde{p}(\lambda, x, v \mid \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [v \tilde{p}(\lambda, x, v \mid \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial v^2} [\sigma^2 v \tilde{p}(\lambda, x, v \mid \cdot)] \\ &\quad + \frac{\partial^2}{\partial x \partial v} [\rho \sigma v \tilde{p}(\lambda, x, v \mid \cdot)] + \lambda_J \int_{\mathbb{R}} \tilde{p}(\lambda, x - j, v \mid \cdot) f_J(j) dj - \lambda_J \tilde{p}(\lambda, x, v \mid \cdot). \end{aligned}$$

As we have from (3.4.9) that $\tilde{p}_S(\lambda, x, v | \cdot) = \frac{f(\lambda)}{\lambda} e^{-\lambda u} \tilde{p}(f(\lambda), x, v)$, replacing λ by $f(\lambda)$ leads to

$$\begin{aligned} f(\lambda) \tilde{p}(f(\lambda), x, v | \cdot) - p(0, x, v | 0, x_s, v_s) &= -\frac{\partial}{\partial x} \left[\left(\mu - \frac{1}{2} v \right) \tilde{p}(f(\lambda), x, v | \cdot) \right] \\ &\quad - \frac{\partial}{\partial v} [\kappa(\gamma - v) \tilde{p}(f(\lambda), x, v | \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [v \tilde{p}(f(\lambda), x, v | \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial v^2} [\sigma^2 v \tilde{p}(f(\lambda), x, v | \cdot)] \\ &\quad + \frac{\partial^2}{\partial x \partial v} [\rho \sigma v \tilde{p}(f(\lambda), x, v | \cdot)] + \lambda_J \int_{\mathbb{R}} \tilde{p}(f(\lambda), x - j, v | \cdot) f_J(j) dj - \lambda_J \tilde{p}(f(\lambda), x, v | \cdot). \end{aligned}$$

Multiplying this last equation by $\frac{f(\lambda)}{\lambda} e^{-\lambda u}$ and using equation (3.4.9), we find

$$\begin{aligned} f(\lambda) \tilde{p}_S(\lambda, x, v | \cdot) - \frac{f(\lambda)}{\lambda} e^{-\lambda u} p(0, x, v | 0, x_s, v_s) &= -\frac{\partial}{\partial x} \left[\left(\mu - \frac{1}{2} v \right) \tilde{p}_S(\lambda, x, v | \cdot) \right] \\ &\quad - \frac{\partial}{\partial v} [\kappa(\gamma - v) \tilde{p}_S(\lambda, x, v | \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [v \tilde{p}_S(\lambda, x, v | \cdot)] + \frac{1}{2} \frac{\partial^2}{\partial v^2} [\sigma^2 v \tilde{p}_S(\lambda, x, v | \cdot)] \\ &\quad + \frac{\partial^2}{\partial x \partial v} [\rho \sigma v \tilde{p}_S(\lambda, x, v | \cdot)] + \lambda_J \int_{\mathbb{R}} \tilde{p}_S(\lambda, x - j, v | \cdot) f_J(j) dj - \lambda_J \tilde{p}_S(\lambda, x, v | \cdot). \end{aligned} \quad (3.4.10)$$

This left-hand term is the Laplace transform $\mathcal{L}$ of the generalized D-C derivative of $p_S(t, x, v | \cdot)$ associated with a Bernstein function f and triplet $(0, 0, \nu)$, as stated in Lemma 3.1. Indeed, since $p(0, x, v | 0, x_s, v_s) = p_S(u, x, v | \cdot)$, we apply Lemma 3.1 and find

$$\mathcal{L} \left[{}^f \mathcal{D}_t^{(u, +\infty)} p_S(t, x, v | \cdot) \right] (\lambda) = f(\lambda) \tilde{p}_S(\lambda, x, v | \cdot) - p(0, x, v | 0, x_s, v_s) \frac{f(\lambda)}{\lambda} e^{-\lambda u},$$

which is precisely the left-hand side of equation (3.4.10). Therefore, applying inverse Laplace transform on both sides of equation (3.4.10), we obtain the announced result (3.4.4) and hence prove Proposition 3.2. $\square$

In the same way as for the non-fractional case in Section 3.2, we can obtain a forward PDE for the Fourier-Laplace transform of the fractional density $p_S(t, x, v | \cdot)$ for the pair (X_{S_t}, V_{S_t}) , which will be a lot easier to solve and whose initial condition does not rely on a Dirac measure. However, before deriving this forward PDE in the bivariate case, it is easier to first analyze the Fourier-Laplace transform of the triplet $(X_{S_t}, V_{S_t}, S_t - S_s)$, which is defined by

$$\begin{aligned} \phi_S(t, z_1, z_2, z_3 | s, w, u, x_s, v_s) & \\ &:= \mathbb{E}^{\mathbb{P}} \left(e^{-iz_1 X_{S_t} - iz_2 V_{S_t} - iz_3 (S_t - S_s)} \mid S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s \right), \end{aligned} \quad (3.4.11)$$

and which satisfies the following proposition.

Proposition 3.3. *For $0 \leq s \leq u < t$, the Fourier-Laplace transform $\phi_S(t, z_1, z_2, z_3 | s, w, u, x_s, v_s)$ of the conditional joint pdf of $(X_{S_t}, V_{S_t}, S_t - S_s)$ satisfies the following*

fractional forward PDE

$$\begin{aligned} {}^f\mathcal{D}_t^{(u,+\infty)}\phi_S(t, z_1, z_2, z_3 | \cdot) &= -z_3\phi_S(t, z_1, z_2, z_3 | \cdot) \\ &\quad - \left(\kappa\gamma z_2 + iz_1\mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi_S(t, z_1, z_2, z_3 | \cdot) \\ &\quad + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - iz_1}{2} \right) \frac{\partial}{\partial z_2} \phi_S(t, z_1, z_2, z_3 | \cdot) \end{aligned} \quad (3.4.12)$$

with initial condition $\phi_S(u, z_1, z_2, z_3 | s, w, u, x_s, v_s) = e^{-iz_1 x_s - z_2 v_s}$ and boundary condition $\phi_S(t, 0, 0, 0 | s, w, u, x_s, v_s) = 1$. For all times t such that $0 \leq s < t \leq u$, then $\phi_S(t, z_1, z_2, z_3 | s, w, u, x_s, v_s) = e^{-iz_1 x_s - z_2 v_s}$.

Proof. The initial and boundary conditions are straightforward given the definition (3.4.11) of $\phi_S(t, z_1, z_2, z_3 | \cdot)$. In order to derive equation (3.4.12), we start by looking at the conditional joint pdf p_S of the triplet $(X_{S_t}, V_{S_t}, S_t - S_s)$ for $0 \leq s \leq u < t$:

$$\begin{aligned} p_S(t, x, v, \tau | s, w, u, x_s, v_s) &:= \\ \frac{\partial^3}{\partial x \partial v \partial \tau} \mathbb{P}(X_{S_t} \leq x, V_{S_t} \leq v, S_t - S_s \leq \tau | S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s) \end{aligned} \quad (3.4.13)$$

From the definition of the conditional pdf $p_S(t, x, v, \tau | \cdot)$ and from equations (3.4.5)-(3.4.6), we have

$$p_S(t, x, v, \tau | s, w, u, x_s, v_s) = p(\tau, x, v | 0, x_s, v_s) g(t, \tau | s, w, u). \quad (3.4.14)$$

From equations (3.3.4) - (3.4.14) and simplifying the notation of the known values $\{s, w, u, x_s, v_s\}$ by $\{\cdot\}$, it follows that

$$\begin{aligned} \tilde{p}_S(\lambda, x, v, \tau | \cdot) &= \int_u^\infty e^{-\lambda t} p_S(t, x, v, \tau | \cdot) dt \\ &= p(\tau, x, v | 0, x_s, v_s) \int_u^\infty e^{-\lambda t} g(t, \tau | s, w, u) dt \\ &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} e^{-\tau f(\lambda)} p(\tau, x, v | 0, x_s, v_s). \end{aligned}$$

We then obtain for the Laplace transform of ϕ_S :

$$\begin{aligned} \tilde{\phi}_S(\lambda, z_1, z_2, z_3 | \cdot) &:= \int_0^{+\infty} \int_0^{+\infty} \int_{-\infty}^{+\infty} \int_u^{+\infty} e^{-\lambda t - iz_1 x - z_2 v - z_3 \tau} p_S(t, x, v, \tau | \cdot) dt dx dv d\tau \\ &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} \int_0^{+\infty} \int_0^{+\infty} \int_{-\infty}^{+\infty} e^{-iz_1 x - z_2 v - z_3 \tau} e^{-\tau f(\lambda)} p(\tau, x, v | 0, x_s, v_s) dx dv d\tau \\ &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} \int_0^{+\infty} e^{-\tau(z_3 + f(\lambda))} \phi(\tau, z_1, z_2 | 0, x_s, v_s) d\tau, \end{aligned} \quad (3.4.15)$$

where $\phi(\tau, z_1, z_2 \mid 0, x_s, v_s) = \int_0^\infty \int_{-\infty}^\infty e^{-iz_1x - z_2v} p(\tau, x, v \mid 0, x_s, v_s) dx dv$. Next, integrating by parts and renaming τ by t yields

$$\begin{aligned} \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) & \\ &= \frac{f(\lambda)}{z_3 + f(\lambda)} e^{-\lambda u} \left[\phi(0, z_1, z_2 \mid 0, x_s, v_s) + \int_0^{+\infty} e^{-t(z_3 + f(\lambda))} \frac{\partial}{\partial t} \phi(t, z_1, z_2 \mid 0, x_s, v_s) dt \right]. \end{aligned} \quad (3.4.16)$$

Proposition (3.2.9) gives the expression of $\frac{\partial}{\partial t} \phi(t, z_1, z_2 \mid 0, x_s, v_s)$ so as to obtain

$$\begin{aligned} \int_0^{+\infty} e^{-t(z_3 + f(\lambda))} \frac{\partial}{\partial t} \phi(t, z_1, z_2 \mid 0, x_s, v_s) dt & \\ &= \int_0^{+\infty} e^{-t(z_3 + f(\lambda))} \left[- \left(\kappa \gamma z_2 + iz_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi(t, z_1, z_2 \mid \cdot) \right. \\ &\quad \left. + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - iz_1}{2} \right) \frac{\partial}{\partial z_2} \phi(t, z_1, z_2 \mid \cdot) \right] dt. \end{aligned} \quad (3.4.17)$$

From this last expression and equation (3.4.15), we find for equation (3.4.16) :

$$\begin{aligned} (z_3 + f(\lambda)) \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) &= \frac{f(\lambda)}{\lambda} e^{-\lambda u} \phi(0, z_1, z_2 \mid 0, x_s, v_s) \\ &\quad - \left(\kappa \gamma z_2 + iz_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) \\ &\quad + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - iz_1}{2} \right) \frac{\partial}{\partial z_2} \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot). \end{aligned}$$

Rearranging the previous equation, we have

$$\begin{aligned} f(\lambda) \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) - \frac{f(\lambda)}{\lambda} e^{-\lambda u} \phi(0, z_1, z_2 \mid 0, x_s, v_s) &= -z_3 \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) \\ &\quad - \left(\kappa \gamma z_2 + iz_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot) \\ &\quad + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - iz_1}{2} \right) \frac{\partial}{\partial z_2} \tilde{\phi}_S(\lambda, z_1, z_2, z_3 \mid \cdot). \end{aligned} \quad (3.4.18)$$

The left-hand side of equation (3.4.18) is the Laplace transform of the generalized D-C derivative of $\phi_S(t, z_1, z_2, z_3 \mid \cdot)$, as given in Lemma 3.1. We have indeed that $\phi_S(u, z_1, z_2, z_3 \mid s, w, u, x_s, v_s) = e^{-iz_1x_s - z_2v_s}$, which is equal to $\phi(0, z_1, z_2 \mid 0, x_s, v_s) = e^{-iz_1x_s - z_2v_s}$. Applying the inverse Laplace transform on both sides leads to the announced result. $\square$

From the previous Proposition 3.3, we can also directly infer the fractional forward PDE satisfied by the Fourier-Laplace transform of the pair (X_{S_t}, V_{S_t}) , which is defined by

$$\begin{aligned} \phi_S(t, z_1, z_2 \mid s, w, u, x_s, v_s) & \\ &:= \mathbb{E}^\mathbb{P} \left[e^{-iz_1X_{S_t} - z_2V_{S_t}} \mid S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s \right]. \end{aligned} \quad (3.4.19)$$

For $0 \leq s \leq u < t$, the Fourier-Laplace transform $\phi_S(t, z_1, z_2 \mid s, w, u, x_s, v_s)$ of the conditional pdf of (X_{S_t}, V_{S_t}) satisfies the following fractional forward PDE :

$$\begin{aligned} {}^f\mathcal{D}_t^{(u, +\infty)} \phi_S(t, z_1, z_2 \mid \cdot) = & - \left(\kappa \gamma z_2 + i z_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi_S(t, z_1, z_2 \mid \cdot) \\ & + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - i z_1}{2} \right) \frac{\partial}{\partial z_2} \phi_S(t, z_1, z_2 \mid \cdot), \end{aligned} \quad (3.4.20)$$

with initial condition $\phi_S(u, z_1, z_2 \mid s, w, u, x_s, v_s) = e^{-i z_1 x_s - z_2 v_s}$ and boundary condition $\phi_S(t, 0, 0 \mid s, w, u, x_s, v_s) = 1$. For all times t such that $0 \leq s < t \leq u$, then $\phi_S(t, z_1, z_2 \mid s, w, u, x_s, v_s) = e^{-i z_1 x_s - z_2 v_s}$.

3.5 Change of measure

We have defined our SSVJ model (3.4.1)-(3.4.2) directly under the physical measure $\mathbb{P}$. Since we want to price European options in illiquid market using this SSVJ model, we now propose a change of measure so as to evaluate our model under an equivalent risk-neutral measure $\mathbb{Q}$. We also show in this section that this SSVJ model is arbitrage-free and leads to an incomplete market. Recall that the subdiffusive asset price process A_S is defined on the filtration $\mathbb{F}^{A \circ S}$. Under $\mathbb{Q}$, we assume in a first step that the stock and variance processes are solutions of the following time-changed stochastic differential equations

$$\frac{dA_{S_t}}{A_{S_t-}} = r dt - \eta \lambda_J m_J dS_t + \sqrt{V_{S_t}} dW_{1,S_t}^{\mathbb{Q}} + (e^{J^{\mathbb{Q}}} - 1) dN_{S_t}^{\mathbb{Q}} \quad (3.5.1)$$

$$dV_{S_t} = \tilde{\kappa} (\tilde{\gamma} - V_{S_t}) dS_t + \sigma \sqrt{V_{S_t}} dW_{2,S_t}^{\mathbb{Q}} \quad (3.5.2)$$

where $W_{j,S}^{\mathbb{Q}}$ ($j = 1, 2$) are time-changed Brownian motions defined on $(\Omega, \mathcal{F}^{A \circ S}, \mathbb{Q})$ such that $\text{Cov}[W_{1,S_t}^{\mathbb{Q}}, W_{2,S_t}^{\mathbb{Q}}] = \rho S_t$. We can therefore rewrite $W_{1,S_t}^{\mathbb{Q}} = \rho W_{2,S_t}^{\mathbb{Q}} + \sqrt{1 - \rho^2} W_{\perp,S_t}^{\mathbb{Q}}$, where $W_{\perp,S}^{\mathbb{Q}}$ is also a time-changed Brownian motion under $\mathbb{Q}$, independent of $W_{2,S}^{\mathbb{Q}}$. Moreover, we denote the $\mathcal{F}_0^{A \circ S}$ -measurable random variable $Y^{\mathbb{Q}} := e^{J^{\mathbb{Q}}} - 1$ and its pdf $f_Y^{\mathbb{Q}}(\cdot)$ under the risk-neutral measure $\mathbb{Q}$ such that $\mathbb{E}^{\mathbb{Q}}[e^{2J^{\mathbb{Q}}}] < \infty$. We further write $m_J = \mathbb{E}^{\mathbb{Q}}[e^{J^{\mathbb{Q}}} - 1] = \mathbb{E}^{\mathbb{Q}}[Y^{\mathbb{Q}}]$ and we have that $N_S^{\mathbb{Q}}$ is a time-changed $\mathbb{Q}$ -Poisson process with intensity $\eta \lambda_J S$. Finally, the risk-free rate r is given by the risk-free bond, whose value at time t is equal to $B_t = e^{rt}$. We will show in what follows the necessity of imposing $r = 0$ in the dynamic (3.5.1).

We now show how to build a change of measure such that the SSVJ model (3.4.1)-(3.4.2) under the physical measure $\mathbb{P}$ in Section 3.4 is equivalent to the above risk-neutral dynamics (3.5.1)-(3.5.2) of this model under $\mathbb{Q}$ (with $r = 0$). Let us first define $\psi(\cdot, \eta) := \ln(\eta f_Y^{\mathbb{Q}}(\cdot) / f_Y^{\mathbb{P}}(\cdot))$ where $\eta \in \mathbb{R}^+$. We consider two $\mathbb{F}^A$ -adapted and predictable processes, denoted $(\theta_{2,t})_{t \geq 0}$ and $(\theta_{\perp,t})_{t \geq 0}$. We next define the time-changed process Z_S as follows

$$Z_{S_t} := \exp \left(-\frac{1}{2} \int_0^{S_t} (\theta_{2,s}^2 + \theta_{\perp,s}^2) ds - \int_0^{S_t} \theta_{2,s} dW_{2,s}^{\mathbb{P}} - \int_0^{S_t} \theta_{\perp,s} dW_{\perp,s}^{\mathbb{P}} + \sum_{k=1}^{N_{S_t}^{\mathbb{P}}} \psi(Y_k^{\mathbb{P}}, \eta) - (\eta - 1)\lambda_J S_t \right), \quad (3.5.3)$$

We first show that the process (3.5.3) is well defined since we consider an a.s. strictly increasing Lévy subordinator U and based on Chapter 10 of Jacod (1979). Indeed, from previous sections, we have that $S = (S_t)_{t \geq 0}$ is a $\mathbb{F}^{A,U}$ -time change with a.s. continuous paths. Hence, any process (and thus in particular $W_2^{\mathbb{P}}, W_{\perp}^{\mathbb{P}}$ and $N^{\mathbb{P}}$) is synchronized with S , see Kobayashi (2011). Therefore, since the (non time-changed) stochastic exponential Z_t exists and is well-defined by Proposition 9.8 of Tankov (2003), the time-changed integral Z_{S_t} given by (3.5.3) also exists and is well-defined by Proposition 10.21 of Jacod (1979). Moreover, as argued in Section 3.4, since Z is a $\mathbb{F}^{A,U}$ -semimartingale and S a finite $\mathbb{F}^{A,U}$ -time change, Z_S is then a $\mathbb{F}^{A \circ S}$ -semimartingale and hence, it admits an equivalent infinitesimal representation by Itô's Lemma for semimartingale with $Y^{\mathbb{P}} := e^{J^{\mathbb{P}}} - 1$

$$dZ_{S_t} = -Z_{S_t-} (\theta_{2,S_t} dW_{2,S_t}^{\mathbb{P}} + \theta_{\perp,S_t} dW_{\perp,S_t}^{\mathbb{P}} - Z_{S_t-} (\eta - 1)\lambda_J dS_t + Z_{S_t-} (e^{\psi(Y^{\mathbb{P}}, \eta)} - 1) dN_{S_t}^{\mathbb{P}}). \quad (3.5.4)$$

3.5.1 Change of measure for the classical SVJ model

We now consider in a first step the non time-changed process Z corresponding to (3.5.3) and we show how a true martingale for Z can be achieved. As in Heston (1993), we first need to impose the additional restriction

$$\theta_{2,t} = \theta_2 \sqrt{V_t}, \quad (3.5.5)$$

where θ_2 is a constant called the market price of volatility risk. A necessary (classical) condition for an equivalent martingale measure to exist is then given by

$$\mu - r + \eta\lambda_J m_J = \sqrt{V_t} \left(\rho\theta_2 \sqrt{V_t} + \sqrt{1 - \rho^2} \theta_{\perp,t} \right), \quad (3.5.6)$$

which ensures that discounted stock prices $(e^{-rt} A_t)_{t \geq 0}$ are local martingales (see equation (3.3) in Wong and Heyde (2006)). Hence, we have for the second market price of risk process

$$\theta_{\perp,t} = \frac{1}{\sqrt{1 - \rho^2}} \left(\frac{\mu - r + \eta\lambda_J m_J}{\sqrt{V_t}} - \rho\theta_2 \sqrt{V_t} \right). \quad (3.5.7)$$

Let us define the measure $\tilde{\mathbb{Q}}$ via the Radon-Nikodym derivative $Z_t = \frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} \Big|_{\mathcal{F}_t^A}$. For

an equivalent martingale measure $\tilde{\mathbb{Q}}$ to exist for A_t , it is required for the process Z to be a true martingale (which is not clear per se due to the unbounded nature of V). Since $\mathbb{E}^{\mathbb{P}}[e^{2J^{\mathbb{P}}}] < \infty$, it is sufficient to verify the martingale property of the corresponding Heston model (without jumps). Hence, we have from Theorem 3.5 of Wong and Heyde (2006) that Z is a true martingale corresponding to the market price of volatility risk defined by (3.5.5) provided that $-\kappa/\sigma \leq \theta_2 < \infty$. It is further shown that the discounted stock price $(e^{-rt}A_t)_{t \geq 0}$ is a $\tilde{\mathbb{Q}}$ -martingale provided that $\kappa + \theta_2\sigma \geq \sigma\rho$ and hence that $\tilde{\mathbb{Q}}$ is an equivalent (risk-neutral) martingale measure for A_t leading to an arbitrage-free market.

3.5.2 Change of measure for the SSVJ model

We now turn to the time-changed process Z_S . Recall that by the First Fundamental Theorem of Asset Pricing, the market model defined by $(\Omega, \mathcal{F}^{A \circ S}, \mathbb{P})$ and asset prices A_S is arbitrage-free if and only if there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that discounted asset prices are martingales with respect to $\mathbb{Q}$, called the equivalent martingale or risk-neutral measure. It is important to note here that discounting must be done using the risk-free asset $B_t^{-1} = e^{-rt}$ but cannot be achieved using the subdiffusive discount factor $B_{S_t}^{-1} = e^{-rS_t}$, leading to a discounted price process given by $(e^{-rt}A_{S_t})_{t \geq 0}$. Indeed, B_{S_t} is asset-specific, while there must be only one risk-free security for each market. Moreover, it is an economic reality that the interest rates accumulate each period in a bank account and hence that the value of the discount factor must change at any time t , ruling out the use of a discount process exhibiting plateaus. A necessary condition similar as equation (3.5.6) for an equivalent martingale measure $\mathbb{Q}$ to exist in the subdiffusive case (3.5.1) is given directly by applying Itô's lemma for semimartingales to $e^{-rt}A_{S_t}$:

$$\left(\mu + \eta \lambda_J m_J - \sqrt{V_{S_t}} \left(\rho \theta_2 \sqrt{V_{S_t}} + \sqrt{1 - \rho^2} \theta_{\perp, S_t} \right) \right) dS_t - r dt = 0, \quad (3.5.8)$$

which ensures that the discounted subdiffusive stock price $(e^{-rt}A_{S_t})_{t \geq 0}$ is a local martingale. This explains the necessity to impose $r = 0$ since the measures dt and dS_t are not equivalent as stochastic processes with respect to $\mathbb{P}$ (dt and dS_t being indeed measures on $\mathbb{R}^+$ whose densities are processes which are not pathwise equivalent for fixed $\omega \in \Omega$, see Proposition 1.8 of Bertoin (1999)).

Since $Z_t = \frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}|_{\mathcal{F}_t^A}$ is a Radon-Nikodym derivative defining an equivalent martingale measure $\tilde{\mathbb{Q}}$ for A_t in Section 3.5.1, we can use Theorem 2 of Torricelli (2020) and Proposition 6.1 of Jacquier and Torricelli (2020) with $r = 0$. We then have that Z_{S_t} is also a Radon-Nikodym derivative defining another equivalent martingale measure $\mathbb{Q}$ for A_{S_t} on $\mathbb{F}^{A \circ S}$ such that $Z_{S_t} = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t^{A \circ S}}$. The next proposition states the dynamic of the Brownian motions and the jump process under this equivalent (martingale) risk-neutral measure $\mathbb{Q}$.

Proposition 3.4. *Under the risk-neutral measure $\mathbb{Q}$ defined by the Radon-Nikodym derivative (3.5.4),*

1. $W_{2,S_t}^{\mathbb{Q}} = W_{2,S_t}^{\mathbb{P}} + \int_0^{S_t} \theta_{2,s} ds$ and $W_{\perp,S_t}^{\mathbb{Q}} = W_{\perp,S_t}^{\mathbb{P}} + \int_0^{S_t} \theta_{\perp,s} ds$ are time-changed Brownian motions with $\theta_{2,t}$ and $\theta_{\perp,t}$ of the form (3.5.5) and (3.5.7), respectively.
2. The process $Y^{\mathbb{Q}} N_S^{\mathbb{Q}}$ is a point process with an intensity equal to $\eta \lambda_J S$ and the pdf of jump is $f_Y^{\mathbb{Q}}(\cdot)$ such that $\mathbb{E}^{\mathbb{Q}}[Y^{\mathbb{Q}}] = m_J$.

Proof. The proof of this proposition is relegated in Appendix 3.A.

Assuming no interest rate in the subdiffusive market as in Magdziarz (2009a) implies that the interest rate accrual is ignored in our model (which is valid for short-term maturities). Another alternative would have been to add directly an absolutely continuous drift $r dt$ in the physical specification (3.4.1) of the SSVJ model, as done *e.g.* in Cartea and Meyer-Brandis (2010), Jacquier and Torricelli (2020) or Torricelli (2020). Note that the market in the SSVJ model is incomplete: the risk-neutral measure $\mathbb{Q}$ is indeed not unique and is related to our choice of dynamics for $\theta_{2,t}$ and $\theta_{\perp,t}$. For the chosen pair $(\theta_{2,t}, \theta_{\perp,t})$ given by (3.5.5) and (3.5.7), the dynamics of A_{S_t} and V_{S_t} under $\mathbb{Q}$ obey respectively the equations

$$\begin{aligned} \frac{dA_{S_t}}{A_{S_t-}} &= -\eta \lambda_J m_J dS_t + \sqrt{V_{S_t}} \left(\rho dW_{2,S_t}^{\mathbb{Q}} + \sqrt{1-\rho^2} dW_{\perp,S_t}^{\mathbb{Q}} \right) + Y^{\mathbb{Q}} dN_{S_t}^{\mathbb{Q}}, \\ dV_{S_t} &= \kappa \left(\gamma - \frac{\sigma \theta_2 V_{S_t}}{\kappa} - V_{S_t} \right) dS_t + \sigma \sqrt{V_{S_t}} dW_{2,S_t}^{\mathbb{Q}}. \end{aligned}$$

We thus obtain that the subdiffusive variance dynamic has the same functional structure under $\mathbb{Q}$ as under the physical measure in (3.4.1) :

$$\frac{dA_{S_t}}{A_{S_t-}} = -\eta \lambda_J m_J dS_t + \sqrt{V_{S_t}} dW_{1,S_t}^{\mathbb{Q}} + Y^{\mathbb{Q}} dN_{S_t}^{\mathbb{Q}}, \quad (3.5.9)$$

$$dV_{S_t} = \tilde{\kappa} (\tilde{\gamma} - V_{S_t}) dS_t + \sigma \sqrt{V_{S_t}} dW_{2,S_t}^{\mathbb{Q}}, \quad (3.5.10)$$

where $\tilde{\kappa}$ and $\tilde{\gamma}$ are the risk-neutral speed and level of mean reversion given by

$$\tilde{\kappa} = \kappa + \theta_2 \sigma, \quad \tilde{\gamma} = \frac{\kappa \gamma}{\kappa + \theta_2 \sigma}. \quad (3.5.11)$$

Notice that since $\theta_2 \geq -\tilde{\kappa}/\sigma$, the mean reversion level $\tilde{\gamma}$ is non-negative under $\mathbb{Q}$. We see that we indeed retrieve the same risk-neutral dynamic as supposed in equations (3.5.1)-(3.5.2) with $r = 0$. Therefore, the change of measure defined by the time-changed Radon-Nikodym derivative (3.5.3) along with the specification (3.5.5)-(3.5.7) allows to work both with the physical and risk-neutral dynamics of the SSVJ model provided that we apply the change of parameters (3.5.11).

3.6 An estimation procedure

In order to estimate the parameters of our model, we consider the daily time series of log-returns from April 2016 to November 2021 for the ALQP stock price, making a total of 1428 data points. Indeed, having no option quotes available in illiquid markets, we are forced to use these log-return data. To the best of our knowledge, such econometric estimation of the parameters of subdiffusive models under the physical probability measure $\mathbb{P}$ is new. We here consider the Particle Markov Chain Monte Carlo (PMCMC) procedure of Andrieu et al. (2010), which is built from the sequential Monte-Carlo filter described in Chapter 2. We therefore work with the physical dynamics (3.4.1)-(3.4.2) of our SSVJ model. Moreover, we need to specify the exact form of the subordinator U and to define the jump size distribution $J^\mathbb{P}$ as well as the counting process $N^\mathbb{P}$ (denoted from now on J and N , respectively). We take here an α -stable process for U , which is well a pure-jump subordinator with no drift satisfying Lemma 3.1 with $f(\lambda) = \lambda^\alpha$. This particular type of subordinator has indeed already been successfully implemented in finance, see *e.g.* Hurst et al. (1999), Carr and Wu (2003) or Cartea and Howison (2009). As in the classical Bates model, we assume J to be normally distributed $J \sim \mathcal{N}(\mu_J, \sigma_J^2)$ and N to be a Poisson process with $N_t \sim Poi(\lambda_J t)$ for all $t \geq 0$. The following estimation procedure can of course be extended easily with other kinds of subordinators (*e.g.* Gamma subordinator) and with other square integrable (Lévy) jump processes.

3.6.1 Filtering

This section introduces the filtering technique that is used to determine the most likely hidden evolution of the inverse subordinator S , the first-passage process U_S , the variance process V_S and the counting process N_S , which are indeed not observed on financial markets (as the only available information are the log-returns). We consider as in Chapter 2 the *Bayesian bootstrap filter* of Gordon et al. (1993), which is a Sequential Importance Resampling (SIR) filter in which the transition prior probability distribution is taken as importance function. This Bayesian bootstrap filter will be combined in the sequel with a Monte-Carlo Markov Chain to estimate the parameters of the SSVJ model but for the moment, we assume that its parameters are known. We consider a discrete sample of log-return observations X_{S_t} and a time step Δt between two observations. The sample is denoted by $x = \{x_1, x_2, \dots, x_n\}$ where $n = T/\Delta t$ (replacing the sample vector r in Algorithm 1). We denote s_j, u_j, v_j and x_j the realizations of $S_j = S_{t_j}, U_{S_j} = U_{S_{t_j}}, V_{S_j} = V_{S_{t_j}}$, and $X_{S_j} = X_{S_{t_j}}$ for $j = 1, \dots, n$ with $t_n = T$. We also denote $\Delta S_j = S_j - S_{j-1}$ and $\Delta N_{S_j} \sim Poi(\lambda_J \Delta S_j)$, with their respective realizations Δs_j and Δn_j . The dynamic of the subdiffusive log-returns is approached by applying Itô's Lemma for semimartingales on equation (3.4.1) and then discretizing the obtained equation,

which gives

$$X_{S_j} = \left(\mu - \frac{V_{S_{j-1}}}{2} \right) \Delta S_j + \rho \sqrt{V_{S_{j-1}}} \Delta W_{2,S_j}^{\mathbb{P}} + \sqrt{1 - \rho^2} \sqrt{V_{S_{j-1}}} \Delta W_{\perp,S_j}^{\mathbb{P}} + J \Delta N_{S_j}, \quad (3.6.1)$$

where $\Delta W_{2,S_j}^{\mathbb{P}} \sim \mathcal{N}(0, \sqrt{\Delta S_j})$, independent from $\Delta W_{\perp,S_j}^{\mathbb{P}} \sim \mathcal{N}(0, \sqrt{\Delta S_j})$. The variance process (3.4.2) is discretized in the same manner by

$$V_{S_j} = V_{S_{j-1}} + \kappa (\gamma - V_{S_{j-1}}) \Delta S_j + \sigma \sqrt{V_{S_{j-1}}} \Delta W_{2,S_j}^{\mathbb{P}}. \quad (3.6.2)$$

The dynamic of S_j is tuned by the parameter α and following Magdziarz (2009b), we can simulate efficiently the sample paths of the discretized version of the inverse α -stable subordinator S_j with

$$S_j = (\min \{k \in \mathbb{N} \mid U_{k\Delta t} > j\Delta t\} - 1) \Delta t. \quad (3.6.3)$$

The α -stable subordinator $U_j = U_{t_j}$ ($j = 1, \dots, n$) is again simulated based on Magdziarz (2009b) using the Euler method and then summing up the increments. We then denote the set of known parameters $\vartheta = (\alpha, \mu, \kappa, \gamma, \sigma, \rho, \mu_J, \sigma_J, \lambda_J)$. The particle filter estimates the most likely sample paths of the hidden inverse subordinator, first-passage, variance and jump processes based on the observed log-returns. First, the information on the realizations $\Delta S_j = \Delta s_j$, $U_{S_j} = u_j$, $V_{S_j} = v_j$, $\Delta N_{S_j} = \Delta n_j$ and $\Delta W_{2,S_j}^{\mathbb{P}} = \Delta w_{2,j}$ is stored in a *particle*, *i.e.* a vector denoted $p_j = (\Delta s_j, u_j, v_j, \Delta n_j, \Delta w_{2,j})$. The visible information (log-returns) up to time t_j is also stored in a vector $x_{1:j} = \{x_1, \dots, x_j\}$ and the density of the observation $f(x_j | p_j)$, conditionally to the information in the particle, is then a Gaussian probability density function. More precisely, we have from equation (3.6.1) that $f(x_j | p_j) = \mathcal{N}(\mu_{r,j}, \sigma_{r,j}^2)$ where

$$\mu_{r,j} = \left(\mu - \frac{v_{j-1}}{2} \right) \Delta s_j + \rho \sqrt{v_{j-1}} \Delta w_{2,j} + \Delta n_j \mu_J,$$

and

$$\sigma_{r,j}^2 = (1 - \rho)^2 v_{j-1} \Delta s_j + \Delta n_j \sigma_J^2.$$

On the other hand, the transition density $f(p_{j+1} | p_j)$ is simulated based on equation (3.6.2), (3.6.3) and on the fact that $\Delta n_j \sim \text{Poi}(\lambda_J \Delta s_j)$. Then, using Bayesian arguments, one can show that the posterior distribution of p_j conditionally to the available information on log-returns at t_j is equal to

$$f(p_j | x_{1:j}) = \frac{f(x_j | p_j)}{\int f(x_j | p_j) f(p_j | x_{1:j-1}) dp_j} f(p_j | x_{1:j-1}), \quad (3.6.4)$$

where we predict

$$f(p_j | x_{1:j-1}) = \int f(p_j | p_{j-1}) f(p_{j-1} | x_{1:j-1}) dp_{j-1}. \quad (3.6.5)$$

Algorithm 3 Particle filtering of the SSVJ model.

Initial step : draw N values of $v_0^{(i)}$ for $i = 1, \dots, N$ from the initial distribution $f(v_0)$, which is assumed to be uniform $\mathcal{U}_{[0,0.36]}$.

Main procedure :

for $j = 1$ to n **do**

Prediction step :

Simulate N times $s_j^{(i)}$ and $u_j^{(i)}$ based on (3.6.3) and compute $\Delta s_j^{(i)} = s_j^{(i)} - s_{j-1}^{(i)}$.

Draw a sample $\Delta w_{2,j}^{(i)}$ from a $\mathcal{N}\left(0, \sqrt{\Delta s_j^{(i)}}\right)$, $i = 1, \dots, N$.

Update $v_j^{(i)}$ using the relation :

$$v_j^{(i)} = v_{j-1}^{(i)} + \kappa \left(\gamma - v_{j-1}^{(i)} \right) \Delta s_j^{(i)} + \sigma \sqrt{v_j^{(i)}} \Delta w_{2,j}^{(i)}$$

Draw N times $\Delta n_j^{(i)}$ from a $Poi(\lambda_J \Delta s_j^{(i)})$.

Correction step :

The particle $p_j^{(i)} = \left(\Delta s_j^{(i)}, u_j^{(i)} v_j^{(i)}, \Delta n_j^{(i)}, \Delta w_{2,j}^{(i)} \right)$ has a probability of occurrence

$$\beta_j^{(i)} = \frac{f(x_j | p_j^{(i)})}{\sum_{i=1:N} f(x_j | p_j^{(i)})}, \quad (3.6.6)$$

where

$$f(x_j | p_j^{(i)}) = \mathcal{N}(\mu_{r,j}, \sigma_{r,j}^2), \quad (3.6.7)$$

with

$$\mu_{r,j} = \left(\mu - \frac{v_{j-1}^{(i)}}{2} \right) \Delta s_j^{(i)} + \rho \sqrt{v_{j-1}^{(i)}} \Delta w_{2,j}^{(i)} + \Delta n_j^{(i)} \mu_J,$$

and

$$\sigma_{r,j}^2 = (1 - \rho)^2 v_{j-1}^{(i)} \Delta s_j^{(i)} + \Delta n_j^{(i)} \sigma_J^2.$$

Calculate the effective sample size, $N_{eff}(j)$:

$$N_{eff}(j) = 1 \left/ \sum_{i=1}^N \left(\beta_j^{(i)} \right)^2 \right. . \quad (3.6.8)$$

Resampling step :

If $N_{eff}(j) < N_{eff,up} = 0.5N$:

Resample with replacement N particles $p_j^{(i)}$ with importance weights $\beta_j^{(i)}$.

The new importance weights are then set to $\beta_j^{(i,2)} = \frac{1}{N}$.

If $N_{eff}(j) > N_{eff,up} = 0.5N$:

No resampling. The new importance weights $\beta_j^{(i,2)}$ are kept as $\beta_j^{(i)}$.

The filtered time scale, first-passage and variance processes at period j are finally computed as the sum of particles weighted by their probability of occurrence:

$$\begin{aligned} \hat{S}_j &= \mathbb{E}^{\mathbb{P}}(S_j | x_{1:j}) = \sum_{i=1}^N s_j^{(i)} \beta_j^{(i,2)}, \quad \hat{U}_{S_j} = \mathbb{E}^{\mathbb{P}}(U_{S_j} | x_{1:j}) = \sum_{i=1}^N u_j^{(i)} \beta_j^{(i,2)}, \\ \hat{V}_{S_j} &= \mathbb{E}^{\mathbb{P}}(V_{S_j} | x_{1:j}) = \sum_{i=1}^N v_j^{(i)} \beta_j^{(i,2)} \end{aligned}$$

end for

Based on the Bayesian bootstrap filter of Gordon et al. (1993), we have that the computation of $f(p_j | x_{1:j})$ can be performed in three steps according to Algorithm 3. First, in the prediction step, we approach $f(p_j | x_{1:j-1})$ by simulations, based on relation (3.6.5). In practice, the integral in (3.6.5) is replaced by a Monte-Carlo simulation of N particles, denoted $p_j^{(i)}$ for $i = 1, \dots, N$. In the correction step, the density $f(p_j | x_{1:j})$ is computed via equation (3.6.4) and is replaced in simulations by the discrete probability of occurrence denoted $\beta_j^{(i)}$ for the particle $i = 1, \dots, N$. We hence have that $\beta_j^{(i)} = \mathbb{P}(p_j^{(i)} | x_{1:j})$, as given by equation (3.6.6). Also note in equation (3.6.7) that if $\Delta s_j^{(i)} = 0$, the pdf $f(x_j | p_j^{(i)})$ is a Dirac measure located at x_j (replaced in practice by a Gaussian pdf with very small variance). In the third step, we perform a resampling with replacement and with probabilities $\beta_j^{(i)}$, $i = 1, \dots, N$, in order to keep track of the most likely sample paths of the different hidden processes. For this resampling step, we first need to calculate an estimate of the effective sample size $N_{eff}(j)$ with equation (3.6.8), as in Doucet et al. (2001). The effective sample size may be equal (or close) to N when the observed process is frozen during a long period of time. In this case, the resampling selects exclusively particles with $\Delta s_j^{(i)} = 0$ which causes degeneracy of the algorithm and which may lead to an over-representation of the same particles after a few iterations. To avoid this, we skip the resampling step if N_{eff} is above $N_{eff,up}$ (typically chosen close to $0.5N$). When a resampling is done, the importance sampling weights are then set to $1/N$. The filtered time scale and first-passage process for the period j are finally computed as the sum of particles weighted by their probability of occurrence

$$\widehat{S}_j = \mathbb{E}^{\mathbb{P}}(S_j | x_{1:j}) = \sum_{i=1}^N s_j^{(i)} \beta_j^{(i,2)} \quad \text{and} \quad \widehat{U}_{S_j} = \mathbb{E}^{\mathbb{P}}(U_{S_j} | x_{1:j}) = \sum_{i=1}^N u_j^{(i)} \beta_j^{(i,2)}, \quad (3.6.9)$$

where $\beta_j^{(i,2)} = 1/N$ in case of resampling and $\beta_j^{(i,2)} = \beta_j^{(i)}$ if no resampling is performed. Similarly, the filtered variance process for the period j is computed as $\widehat{V}_{S_j} = \mathbb{E}^{\mathbb{P}}(V_{S_j} | x_{1:j}) = \sum_{i=1}^N v_j^{(i)} \beta_j^{(i,2)}$, where $\beta_j^{(i,2)}$ is defined as in (3.6.9). Following Doucet et al. (2001), the log-likelihood of the whole sample path is approached before resampling by the sum

$$\begin{aligned} \ln f(x | \vartheta) &= \sum_{j=1}^n \ln f(x_j | x_{j-1}) = \sum_{j=1}^n \ln \left(\int f(x_j | p_j) f(p_j | x_{j-1}) dp_j \right) \\ &= \sum_{j=1}^n \ln \left(\sum_{i=1}^N f(x_j | p_j^{(i)}) \beta_{j-1}^{(i,2)} \right). \end{aligned} \quad (3.6.10)$$

To illustrate this section on filtering, we simulate via (3.6.1),(3.6.2),(3.6.3) a sample path of A_{S_t} (and hence also of S_t , V_{S_t} , N_{S_t} and U_{S_t}). Under the assumption that

$$\begin{aligned} \alpha &= 0.8, \quad \mu = 0.05, \quad \kappa = 0.8, \quad \gamma = 0.15, \quad \sigma = 0.9, \\ \rho &= -0.6, \quad \mu_J = -0.2, \quad \sigma_J = 0.3, \quad \lambda_J = 0.9, \end{aligned} \quad (3.6.11)$$

we consider 1260 observations of A_{S_t} over five years of trading (time step $\Delta t = \frac{1}{252}$). The particle filter runs with 10 000 particles. Figure 3.6.1 compares the simulated sample paths of S_t and V_{S_t} to their filtered estimate obtained solely from the simulated values of A_{S_t} via Algorithm 3. This confirms the ability of the filter to estimate the hidden sample paths of the stochastic clock and the subordinated variance process based exclusively on the observed log-returns of the illiquid asset.

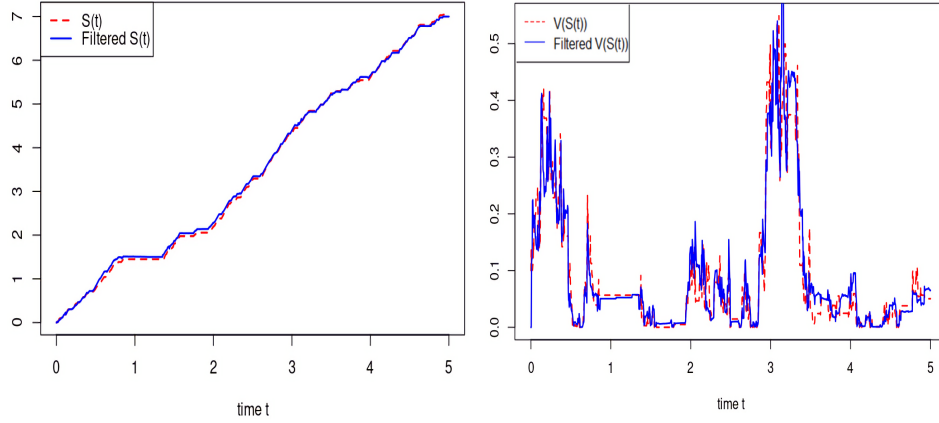


Figure 3.6.1: The figure depicts in red dotted line the simulated inverse subordinator S (left plot) and variance process V_S (right plot) based on parameters (3.6.11), compared with their filtered estimate in blue obtained from the path of A_S via Algorithm 3.

This Algorithm 3 is also used to estimate the first-passage process U_S from the illiquid asset price, as depicted in Figure 3.6.2. We can see that at price moves, the estimation is perfect since in this case $U_{S_j} = t_j$. However, when the price is frozen at time t_j , U_{S_j} is equal to the last time $t_{j'} > t_j$ at which the price is frozen and for which $S_{j'} = S_j$, which is of course an unobserved event at time t_j (see Section 3.3 on filtrations). Our estimation procedure provides in this case the best estimate of U_{S_j} based on the estimated values of $(S_0, S_1, \dots, S_j)$ up to time t_j . More precisely, Algorithm 3 computes from equation (3.6.9) the average behavior of U_{S_j} knowing all the values of S_j up to time t_j as we need the information on the length of time already spent by the process S (and hence by A_S) in the frozen state since its last move. Indeed, recall that U_S is non-Markovian (U_S is in fact not $\mathbb{F}^S$ -adapted) and hence, we have $\mathbb{E}^\mathbb{P}[U_{S_j} | S_0, \dots, S_j] \neq \mathbb{E}^\mathbb{P}[U_{S_j} | S_j]$. We can then clearly see on Figure 3.6.2 that the exact behavior of U_{S_j} ($j = 1, \dots, n$) cannot be reproduced by the estimation procedure (again due to its unobserved nature at time t_j) but the best that can be achieved when the price is frozen is to compute its average behavior conditionally on the estimation of $\widehat{S}_0, \dots, \widehat{S}_j$, as depicted on the plot. This is why, we tend to underestimate U_S during long motionless peri-

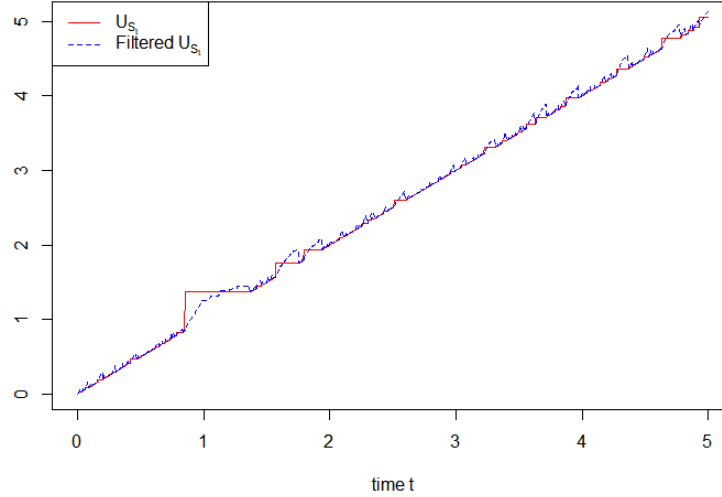


Figure 3.6.2: In red, simulated sample path of the first-passage process U_S based on the parameters (3.6.11) and its filtered estimate $\hat{U}_S$ in blue dotted line.

ods and to overestimate it for short motionless periods. Moreover, we clearly see in Figures 3.6.2 and 3.B.1 that the estimation of U_{S_j} is not constant even when S_j is frozen, which results from the fact that information about U_S accumulate through time since the estimation procedure takes into account the information on the (increasing) length of time since the last move of S . These behaviors are confirmed in Appendix 3.B with a second simulated sample path of A_S using the same parameters (3.6.11). This estimation procedure hence allows to consider a more general setting than Theorem 3 of Cartea and Meyer-Brandis (2010) in which valuation can only be effected at the exact times of price moves.

This filtering technique based on a bootstrap SIR filter has already been successfully implemented in finance as in Johannes and Polson (2010), see also Chapter 2 of this thesis. Other filters exist for filtering SVJ models, such as the Auxiliary Particle Filtering (APF) of Johannes et al. (2009). However, for the Bayesian estimation of SVJ models, Jacobs and Liu (2018) find that the APF yields more variable likelihood evaluations and is therefore less suitable than the Bayesian bootstrap filter.

3.6.2 Econometric estimation

We have so far considered that the set of parameters ϑ is known, which is not the case in practice. An inherent problem of particle filters is that the estimate

of the likelihood is not a smooth function of parameters. Practically, calibrating the model by maximizing the log-likelihood (*e.g.* with gradient descent) is then uncertain. We here prefer using the PMCMC procedure which has been shown by Johannes and Polson (2010) to be highly efficient for such financial estimation of stochastic volatility models. We now denote the set of unknown parameters by $\Theta = (\alpha, \mu, \kappa, \gamma, \sigma, \rho, \mu_J, \sigma_J, \lambda_J)$. We here again adopt a Bayesian approach and Θ is a multivariate random variable with realizations ϑ and with state space $\mathcal{X}$. The parameters posterior distribution, based on the observed sample $x = \{x_1, \dots, x_n\}$ is

$$f(\vartheta | x) = \frac{f(\vartheta)f(x | \vartheta)}{\int_{\mathcal{X}} f(\vartheta')f(x | \vartheta') d\vartheta'} , \quad (3.6.12)$$

where $f(\vartheta)$ and $f(x | \vartheta)$ denote respectively the parameter prior distribution and the likelihood of the data approached by the estimate (3.6.10). The density $f(\vartheta | x)$ is built by the PMCMC method, as described previously in Section 2.4.2 and Algorithm 2 (and replacing r with x).

We first apply this PMCMC algorithm on the simulated sample path A_{S_t} above to see if we can obtain a consistent estimation of the true parameters (3.6.11) used to generated this sample path. The standard deviation σ_{ϑ} of the transition distribution is set to

$$\sigma_{\vartheta} = (0.1\%, 0.5\%, 0.6\%, 0.1\%, 0.6\%, 0.5\%, 0.2\%, 0.1\%, 0.2\%)$$

Even if the procedure converges theoretically to parameter estimates, it is advisable to provide realistic initial parameters to improve the speed of convergence. Therefore, we impose $\vartheta_0 = (0.7, 0.01, 0.2, 0.1, 0.1, -0.2, 0, 0.1, 0.7)$. The filter runs with $N = 10000$ particles and we perform $K = 1000$ iterations of the PMCMC procedure. We hence obtain Table 3.6.1 with a burn-in period of 500 iterations.

Table 3.6.1: Parameter estimates of the SSVJ model based on the simulated sample path of A_{S_t} from the parameters (3.6.11). The second line of the table gives the standard error of these estimates.

	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\mu}_J$	$\hat{\sigma}_J$	$\hat{\lambda}_J$
PMCMC estimates	0.782	0.051	0.495	0.194	0.862	-0.531	-0.283	0.282	1.087
PMCMC std. err.	0.041	0.017	0.173	0.040	0.077	0.115	0.029	0.018	0.044

The estimated parameters of Table 3.6.1 are quite close to the true ones (3.6.11), except for κ which is typically known to have a higher standard deviation in such SVJ models and hence to be harder to estimate from the PMCMC procedure (see *e.g.* Table 2.5.1 in Chapter 2). Moreover, the obtained log-likelihood (3.6.10) is here equal to 15 141.73, which is very close to the one derived with the true parameters equal to 15 142.12. This confirms the overall good quality of the fit when using this

PMCMC procedure with a Bayesian bootstrap filter.

We now use the PMCMC algorithm to estimate the set of parameters Θ of the SSVJ model (3.4.1)-(3.4.2), based on the observed time series of log-returns for the ALQP stock price between April 2016 and November 2021. The same settings $(\sigma_\vartheta, \theta_0, N, K)$ as in the above simulated example are again used in Algorithm 3, leading to Table 3.6.2 with the estimated parameters for ALQP as well as their associated standard error, and with a log-likelihood (3.6.10) equal to 25 059.34. Note that the log-likelihood obtained for the ALQP stock price with the standard SVJ model is equal to 5035.17, which proves that the SSVJ model is a much more consistent and accurate model in illiquid markets.

Table 3.6.2: Parameter estimates of the SSVJ model for the ALQP stock price and standard error of these estimates.

	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\mu}_J$	$\hat{\sigma}_J$	$\hat{\lambda}_J$
PMCMC estimates	0.581	-0.021	0.237	0.050	0.309	-0.806	-0.056	0.037	0.947
PMCMC std. err.	0.013	0.019	0.127	0.024	0.018	0.066	0.021	0.031	0.076

We see that we have a quite low value of $\hat{\alpha}$ coming from the fact that we observe relatively long periods of constant values in the quotes of the ALQP stock price in Figure 3.1.1. We also have a negative mean return with on average negative jumps. The estimated values related to the variance process are quite standard with a long-term volatility of $\sqrt{0.05} = 22.3\%$ and with a significant leverage effect since $\rho = -0.81 < 0$. Finally, note that the conditions of Section 3.5.1 for an equivalent martingale measure to exist are verified with $\theta_2 = 0$.

3.7 Risk management and option pricing

Recall that the estimation procedure was performed using an α -stable subordinator U . The Lévy triplet of U satisfies in this case Lemma 3.1 since it is given by $(0, 0, \nu)$ where $\nu(dz) = \frac{\alpha}{\Gamma(1-\alpha)} \frac{dz}{z^{1+\alpha}}$. The Lévy tail measure then satisfies

$$\bar{\nu}(s)ds = ds \int_s^\infty \frac{\alpha z^{-\alpha-1}}{\Gamma(1-\alpha)} dz = \frac{s^{-\alpha}}{\Gamma(1-\alpha)} ds.$$

The D-C derivative of $h(t)$ becomes in this case the Caputo fractional derivative, denoted ${}_u D^\alpha h(t)$. Indeed, if we perform the change of variable $s' = t - s$, we infer via equation (3.3.8) that

$$\int_0^{t-u} \frac{\partial}{\partial t} h(t-s) \bar{\nu}(s) ds = \int_u^t \frac{\partial}{\partial s'} h(s') \bar{\nu}(t-s') ds',$$

and we thus obtain the following Caputo fractional derivative as defined in Ishteva et al. (2005) :

$${}^f\mathcal{D}_t^{(u,+\infty)}h(t) = {}_uD^\alpha h(t) := \frac{1}{\Gamma(1-\alpha)} \int_u^t \frac{h'(s)}{(t-s)^\alpha} ds. \quad (3.7.1)$$

In the case of the α -stable subordinator, the general fractional forward PDE (3.4.20) satisfied by the Fourier-Laplace transform of the pdf $p_S(t, x, v | \cdot)$ associated with the bivariate process (X_S, V_S) hence becomes the following fractional Caputo PDE for $t > u \geq s \geq 0$:

$$\begin{aligned} {}_uD^\alpha \phi_S(t, z_1, z_2 | \cdot) = & - \left(\kappa \gamma z_2 + i z_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi_S(t, z_1, z_2, z_3 | \cdot) \\ & + \left(-z_2(\kappa + i\rho\sigma z_1) - \frac{\sigma^2 z_2^2}{2} + \frac{z_1^2 - i z_1}{2} \right) \frac{\partial}{\partial z_2} \phi_S(t, z_1, z_2, z_3 | \cdot), \end{aligned} \quad (3.7.2)$$

with the same initial condition $\phi_S(u, z_1, z_2 | \cdot) = e^{-iz_1 x_s - z_2 v_s}$ and boundary condition $\phi_S(t, 0, 0 | \cdot) = 1$. Since we consider a Gaussian density for modeling the jump size J in our setting, we obtain the following Fourier transform

$$\hat{f}_J(z_1) = \int_{-\infty}^{+\infty} e^{-iz_1 j} f_J(j) dj = e^{-i\mu_J z_1 - \frac{1}{2} z_1^2 \sigma_J^2}, \text{ where } f_J(j) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{j - \mu_J}{\sigma} \right)^2}.$$

We now propose a numerical method for solving the fractional Caputo PDE (3.7.2). Since for risk management in Section 3.7.1 and option pricing in Section 3.7.2 (second alternative), we are only interested in approximating the Fourier-Laplace transform $\phi_S(t, z_1, z_2 | \cdot)$ at $z_2 = 0$, our numerical method is based on a series expansion of $\phi_S(t, z_1, z_2 | \cdot)$ in the neighborhood of $z_2 = 0$. This approximation is standard in physics for non-fractional PDE, see *e.g.* Bildik et al. (2006). Moreover, we know from the generalized Taylor's formula of Odibat and Shawagfeh (2007) that any $\mathcal{C}^\infty$ function $l(t, z)$ with respect to time t may be expressed as the infinite sum

$$l(t, z) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k\alpha + 1)} \left[\left(\frac{\partial}{\partial t^\alpha} \right)^k l(t, z) \right]_{t=t_0} (t - t_0)^{k\alpha}. \quad (3.7.3)$$

We understand $\left(\frac{\partial}{\partial t^\alpha} \right)^k$ as the k^{th} order Caputo derivative (*i.e.* applying k times the Caputo derivative). We refer to equation (3.19) of Odibat and Shawagfeh (2007) where the exact expression for the radius of convergence R of the generalized Taylor series (3.7.3) is given. Moreover, if we assume $l(t, z)$ is the product of two $\mathcal{C}^\infty$ functions $f(t)$ and $g(z)$, a Taylor's development of $g(z)$ around z_0 leads to the following expansion for $l(t, z)$

$$\begin{aligned} l(t, z) &= \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} \frac{1}{\Gamma(k\alpha + 1)} \frac{1}{h!} \left[\left(\frac{\partial}{\partial t^\alpha} \right)^k f(t) \right]_{t=t_0} \left[\frac{\partial^h g(z)}{\partial z^h} \right]_{z=z_0} (t - t_0)^{k\alpha} (z - z_0)^h \\ &= \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} w_\alpha(k, h) (t - t_0)^{k\alpha} (z - z_0)^h, \end{aligned} \quad (3.7.4)$$

where $(w_\alpha(k, h))_{k, h > 0} := \frac{1}{\Gamma(k\alpha + 1)} \frac{1}{h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} l(t, z) \right]_{t=t_0, z=z_0}$. We here use the notation

$$\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} := \left(\frac{\partial}{\partial t^\alpha} \right)^k \frac{\partial^h}{\partial z^h},$$

but the reader must keep in mind that $\left(\frac{\partial}{\partial t^\alpha} \right)^k \neq \frac{\partial^k}{\partial t^{k\alpha}}$ in the case of the Caputo derivative. Then, for a fixed value of z_1 , the Fourier-Laplace transform $\phi_S(t, z_1, z_2 | \cdot)$ is a function of z_2 with $t > u$. We rewrite here this transform $\phi_S(t, z_1, z | \cdot)$ to emphasize that we consider it as a function of $z = z_2$ and we can then approximate⁴ $\phi_S(t, z_1, z | \cdot)$ by the following expansion around $z = 0$:

$$\phi_S(t, z_1, z | s, w, u, x_s, v_s) \approx \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} w_\alpha(k, h) \times z^h (t - u)^{\alpha k}, \quad (3.7.5)$$

where the weights in this sum are obtained as above for $h > 0$ by

$$w_\alpha(k, h) = \frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \phi_S(t, z_1, z | s, w, u, x_s, v_s) \right]_{t=u, z=0}. \quad (3.7.6)$$

The weights are then computed recursively thanks to the following proposition.

Proposition 3.5. *The differential weights (3.7.6) verify the following relations*

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \frac{\partial^\alpha \phi_S(t, z_1, z | \cdot)}{\partial t^\alpha} \right]_{t=u, z=0} = \frac{\Gamma((k+1)\alpha + 1)}{\Gamma(k\alpha + 1)} w_\alpha(k+1, h) \quad (3.7.7)$$

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z \phi_S(t, z_1, z | \cdot) \right]_{t=u, z=0} = w_\alpha(k, h-1) \quad (3.7.8)$$

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \frac{\partial \phi_S(t, z_1, z | \cdot)}{\partial z} \right]_{t=u, z=0} = (h+1) w_\alpha(k, h+1) \quad (3.7.9)$$

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z \frac{\partial \phi_S(t, z_1, z | \cdot)}{\partial z} \right]_{t=u, z=0} = h w_\alpha(k, h) \quad (3.7.10)$$

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z^2 \frac{\partial \phi_S(t, z_1, z | \cdot)}{\partial z} \right]_{t=u, z=0} = (h-1) w_\alpha(k, h-1) \quad (3.7.11)$$

⁴The approximation (3.7.5) is exact if $\phi_S(t, z_1, z | \cdot)$ is the product of one function of t and one function of z .

Proof. Equations (3.7.7) and (3.7.9) directly follow from the definition of the differential weights. The relation (3.7.8) is a straightforward result from the following equality with $z = 0$:

$$\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z \phi_S(t, z_1, z | \cdot) = h \frac{\partial^{k\alpha+h-1}}{\partial t^{k\alpha} \partial z^{h-1}} \phi_S(t, z_1, z | \cdot) + z \frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \phi_S(t, z_1, z | \cdot),$$

whereas equation (3.7.10) comes from

$$\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z \frac{\partial \phi_S(t, z_1, z | \cdot)}{\partial z} = h \frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \phi_S(t, z_1, z | \cdot) + z \frac{\partial^{k\alpha+h+1}}{\partial t^{k\alpha} \partial z^{h+1}} \phi_S(t, z_1, z | \cdot),$$

and equation (3.7.11) from

$$\begin{aligned} \frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} z^2 \frac{\partial \phi_S(t, z_1, z | \cdot)}{\partial z} &= h(h-1) \frac{\partial^{k\alpha+h-1}}{\partial t^{k\alpha} \partial z^{h-1}} \phi_S(t, z_1, z | \cdot) \\ &+ 2hz \frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \phi_S(t, z_1, z | \cdot) + z^2 \frac{\partial^{k\alpha+h+1}}{\partial t^{k\alpha} \partial z^{h+1}} \phi_S(t, z_1, z | \cdot). \end{aligned}$$

□

The next proposition hence provides an approximation of the Fourier-Laplace transform $\phi_S(t, z_1, z | \cdot)$ by truncating the asymptotic expansion (3.7.5).

Proposition 3.6. *The Fourier-Laplace transform of the bivariate process (X_{S_t}, V_{S_t}) at $t > u \geq s \geq 0$ is approximated around $z = 0$ by the following truncated sum*

$$\phi_S(t, z_1, z | s, w, u, x_s, v_s) \approx \sum_{k=0}^K \sum_{h=0}^H w_\alpha(k, h) \times z^h (t-u)^{\alpha k}, \quad (3.7.12)$$

where the differential weights $w_\alpha(k, h)$ verify the following recursive equation for $h > 0$

$$\begin{aligned} \frac{\Gamma((k+1)\alpha+1)}{\Gamma(k\alpha+1)} w_\alpha(k+1, h) &= -\kappa\gamma w_\alpha(k, h-1) \\ &- \left(iz_1\mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) w_\alpha(k, h) - (\kappa + i\rho\sigma z_1) h w_\alpha(k, h) \\ &- \frac{\sigma^2}{2} (h-1) w_\alpha(k, h-1) + (h+1) \frac{(z_1^2 - iz_1)}{2} w_\alpha(k, h+1). \end{aligned} \quad (3.7.13)$$

For $h = 0$, the weights are obtained via

$$\frac{\Gamma((k+1)\alpha+1)}{\Gamma(k\alpha+1)} w_\alpha(k+1, 0) = - \left(iz_1\mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) w_\alpha(k, 0) + \frac{(z_1^2 - iz_1)}{2} w_\alpha(k, 1).$$

The initial condition on the weights for initializing the recursion is given by

$$w_\alpha(0, h) = e^{-iz_1 x_s} \frac{1}{h} (-v_s)^h.$$

Proof. To establish the recursion (3.7.13) for $h > 0$, we apply the differential operator

$$\frac{1}{\Gamma(k\alpha + 1)h!} \left[\frac{\partial^{k\alpha+h}}{\partial t^{k\alpha} \partial z^h} \right]_{t=u, z=0},$$

on the forward Caputo PDE (3.7.2). Next, for $h = 0$, we first notice from approximation (3.7.12) that

$$\phi_S(t, z_1, 0 | \cdot) \approx \sum_{k=0}^K \sum_{h=0}^H w_\alpha(k, h) \times 0^h (t-u)^{\alpha k} = \sum_{k=0}^K w_\alpha(k, 0) \times (t-u)^{\alpha k}.$$

Therefore, we see from equation (3.7.3) that the weight in the case $h = 0$ is given by

$$w_\alpha(k, 0) = \frac{1}{\Gamma(k\alpha + 1)} \left[\frac{\partial^{k\alpha}}{\partial t^{k\alpha}} \phi_S(t, z_1, 0 | \cdot) \right]_{t=u}. \quad (3.7.14)$$

Then, similarly as in Proposition 3.5, we find the following relation

$$\frac{1}{\Gamma(k\alpha + 1)} \left[\frac{\partial^{k\alpha}}{\partial t^{k\alpha}} \frac{\partial \phi_S(t, z_1, 0 | \cdot)}{\partial z} \right]_{t=u} = w_\alpha(k, 1),$$

and applying the operator $\frac{1}{\Gamma(k\alpha + 1)} \left[\frac{\partial^{k\alpha}}{\partial t^{k\alpha}} \right]_{t=u}$ on the following Caputo PDE with $z_2 = 0$,

$$\begin{aligned} {}_u D^\alpha \phi_S(t, z_1, 0 | \cdot) = \\ - \left(iz_1 \mu - \lambda_J \left(\hat{f}_J(z_1) - 1 \right) \right) \phi_S(t, z_1, 0 | \cdot) + \frac{z_1^2 - iz_1}{2} \frac{\partial}{\partial z} \phi_S(t, z_1, 0 | \cdot), \end{aligned}$$

leads to the announced result. Finally, since

$$\phi_S(u, z_1, z | \cdot) \approx \sum_{k=0}^K \sum_{h=0}^H w_\alpha(k, h) \times z^h 0^{\alpha k} = \sum_{h=0}^H w_\alpha(0, h) z^h,$$

and since

$$\phi_S(u, z_1, z | \cdot) = e^{-iz_1 x_s - z v_s} = e^{-iz_1 x_s} \sum_{h=0}^{\infty} \frac{1}{h!} (-v_s)^h z^h,$$

we find $w_\alpha(0, h) = e^{-iz_1 x_s} \frac{1}{h!} (-v_s)^h$ for the initial weight of the recursion. $\square$

3.7.1 Risk management

We first use the series expansion of the Fourier-Laplace transform derived in Proposition 3.6 to compute the moments of the subdiffusive stock price A_{S_T} at maturity T (provided that they exist), as well as some risk measures related to A_{S_T} under the SSVJ model. The existence of the first four unconditional central moments of A_{S_T} at some maturity T is detailed in Appendix 3.C. The probability distribution of the subdiffusive asset price A_{S_T} can then be obtained easily via its Fourier representation. By inverting $\phi_S(T, z_1, 0 | \cdot)$, it is indeed possible to determine the density of probability $p_S(T, x | \cdot) = \int_{\mathbb{R}^+} p_S(T, x, v | \cdot) dv$ of the log-returns X_{S_T} and then compute numerically the moments of order k of the subdiffusive asset price via $\mathbb{E}^\mathbb{P}[A_{S_T}^k | \cdot] = \int_{\mathbb{R}} e^{kx} p_S(T, x | \cdot) dx$, as well as the related Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR). More precisely, we consider here the financial portfolio Y consisting of one long position in the subdiffusive stock up to the initial selling price, *i.e.* $Y_{S_T} := A_{S_T} - A_0$. Following Artzner et al. (1999), we define in this chapter⁵ the VaR at level ε of this portfolio as $VaR_\varepsilon(Y_{S_T}) := \inf\{q \in \mathbb{R} \mid \mathbb{P}(Y_{S_T} + q < 0) \leq \varepsilon\}$, which is the smallest amount of capital which, if added to Y_{S_T} and invested in the risk-free asset, keeps the probability of a negative outcome (loss) below the level ε . Similarly, the TVaR at level ε for this long position is defined by $TVaR_\varepsilon(Y_{S_T}) := \mathbb{E}^\mathbb{P}[-Y_{S_T} | Y_{S_T} \leq -VaR_\varepsilon(Y_{S_T})]$, which is the expected loss given that this loss exceeds the VaR at level ε . For assessing the risk of a long position, we are typically interested in small values of ε (0.5%, 1%). The inversion of the Fourier transform is then usually performed accurately and fast using a Discrete Fast Fourier Transform (DFFT) algorithm, see Appendix 2.F in Chapter 2. With the parameters $(\hat{\alpha}, \hat{\mu}, \hat{\kappa}, \hat{\gamma}, \hat{\sigma}, \hat{\rho}, \hat{\mu}_J, \hat{\sigma}_J, \hat{\lambda}_J)$ estimated from the PMCMC algorithm in Table 3.6.2, we obtain in Table 3.7.1 the moments and related risk measures of the ALQP share price A_{S_T} , from $s = 0$ with $T = 0.5$ years and with $A_0 = 24.8$ ($x_0 = \ln 24.8$). Notice from Appendix 3.C that the first four unconditional moments of A_{S_T} are finite for $T = 0.5$. A comparison with the classical SVJ model based on the same parameters (and $\alpha = 1$) is also performed and the density of A_{S_T} is depicted under the two models in Figure 3.7.1.

Table 3.7.1: The first columns give the first four central moments of A_{S_T} under the classical SVJ model and the SSVJ model. The four last columns give the Value-at-Risk and Tail Value-at-Risk of the related portfolio Y_{S_T} at levels $\varepsilon = 0.5\%$ and 1% .

	Expect.	Variance	Skew.	Kurt.	$VaR_{0.5\%}$	$VaR_{1\%}$	$TVaR_{0.5\%}$	$TVaR_{1\%}$
SVJ	24.543	7.256	-0.317	4.039	8.513	7.554	9.770	8.935
SSVJ	24.401	11.566	-0.187	7.637	12.264	10.189	15.490	13.296

⁵This definition did not make sense for Chapter 2 as the CPPI portfolio P could not be below the (non-negative) floor F .

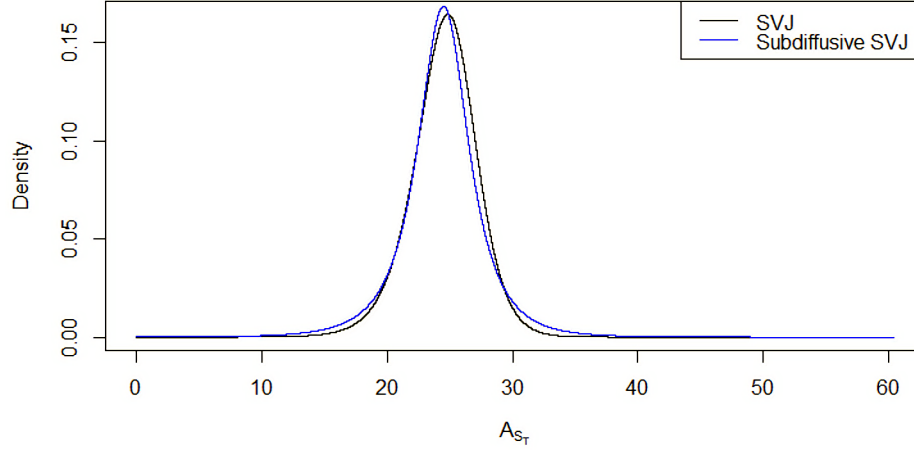


Figure 3.7.1: Density of the ALQP stock price (A_{S_T}) at $T = 0.5$ with $s = 0$ and $x_0 = \ln(24.8)$ under the classical SVJ model (black line) and the SSVJ model (blue line).

We clearly see from Table 3.7.1 and Figure 3.7.1 that the SSVJ model leads to a higher variability of A_{S_T} and to fatter left and right tails, making this model more risky. This is confirmed by the VaR and TVaR at levels 0.5% and 1% for Y_{S_T} .

3.7.2 Pricing

We now illustrate how to use our SSVJ model for pricing options in illiquid markets. More precisely, we start by showing how to price a European call with maturity T and strike K , whose value at time s under the risk-neutral SSVJ model (3.5.9)-(3.5.10) is given by

$$\begin{aligned} C_S(T, K, S_s, U_{S_s}, A_{S_s}, V_{S_s}) &= \mathbb{E}^{\mathbb{Q}} \left[e^{-r(T-s)} (A_{S_T} - K)_+ \middle| S_s, U_{S_s}, A_{S_s}, V_{S_s} \right] \\ &= \mathbb{E}^{\mathbb{Q}} \left[e^{-r(T-s)} (e^{X_{S_T}} - K)_+ \middle| S_s, U_{S_s}, e^{X_{S_s}}, V_{S_s} \right], \end{aligned}$$

with $X_{S_s} = \ln A_{S_s}$. In our context, conditionally to the information about the stochastic clock $S_s = w$, the first-passage $U_{S_s} = u$, the stock value $A_{S_s} = e^{x_s}$ and the variance $V_{S_s} = v_s$, we can write

$$C_S(T, K, w, u, x_s, v_s) = \mathbb{E}^{\mathbb{Q}} \left[(e^{X_{S_T}} - K)_+ \middle| S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s \right]. \quad (3.7.15)$$

since we consider a zero risk-free rate $r = 0$ in this chapter. Using the conditional pdf of X_{S_t} , denoted by $p_S(t, x \mid s, w, u, x_s, v_s) = \int_0^\infty p_S(t, x, v \mid s, w, u, x_s, v_s) dv$,

we can rewrite the price at time s of the European call by

$$C_S(T, K, w, u, x_s, v_s) = \int_{-\infty}^{\infty} (e^x - K)_+ p_S(T, x \mid s, w, u, x_s, v_s) dx. \quad (3.7.16)$$

From this point, we can price numerically this European call using several methods. The most direct one is based on classical Monte-Carlo simulations. However, these simulations are computer-intensive and we hence rather consider the two following alternatives.

- First Alternative :

From the independence between S and X , we can use equation (3.4.8) to rewrite the price (3.7.16) by

$$\begin{aligned} C_S(T, K, w, u, x_s, v_s) &= \int_0^\infty \int_{-\infty}^\infty (e^x - K)_+ p(\tau, x \mid 0, x_s, v_s) g(T, \tau \mid s, w, u) dx d\tau \\ &= \int_0^\infty C(\tau, K, x_s, v_s) g(T, \tau \mid s, w, u) d\tau, \end{aligned} \quad (3.7.17)$$

where $C(\tau, K, x_s, v_s)$ is the call price with maturity τ under the risk-neutral dynamic of the classical SVJ model with $r = 0$, which is given by

$$C(\tau, K, x_s, v_s) = \mathbb{E}^\mathbb{Q} [(A_\tau - K)_+ \mid A_0 = e^{x_s}, V_0 = v_s].$$

This European call under the SVJ model is traditionally computed in a very efficient way using Fourier methods, as in Gatheral (2011). This method works well when the conditional density $g(T, \tau \mid \cdot)$ of the inverse subordinator can be obtained easily, which is the case with the α -stable subordinator U . Indeed, we have from Hainaut (2021) that the probability density function of $S_T - S_s$ conditionally to $S_s = w$ and $U_{S_s} = u$, can be represented as an infinite sum

$$g(T, \tau \mid s, w, u) = \frac{1}{\alpha\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\Gamma(\alpha k + 1) \sin(\pi k \alpha)}{k!} \frac{\tau^{k-1}}{(T - u)^{\alpha k}}. \quad (3.7.18)$$

However, when the conditional density $g(T, \tau \mid \cdot)$ is hard to derive, we rather use a second alternative.

- Second Alternative :

We now consider the call price (3.7.16) as a function of the log-strike $k := \ln K$ so as to use the methodology of Carr and Madan (1999). Since this function is not integrable, its Fourier transform is not defined. However, if we fix $\varepsilon > 1$, then the function

$$c_S(k) := e^{\varepsilon k} \mathbb{E}^\mathbb{Q} \left[(e^{X_{S_T}} - e^k)_+ \mid S_s = w, U_{S_s} = u, X_{S_s} = x_s, V_{S_s} = v_s \right],$$

is square integrable and its Fourier transform exists. We denote it by $\hat{c}_S(\omega) := \int_{-\infty}^{+\infty} e^{-i\omega k} c_S(k) dk$. Using Fubini's theorem for non-negative functions with Lebesgue measure, we then have

$$\begin{aligned}
\hat{c}_S(\omega) &= \int_{-\infty}^{\infty} e^{(\varepsilon - i\omega)k} C_S(T, e^k, w, u, x_s, v_s) dk \\
&= \int_{-\infty}^{\infty} e^{(\varepsilon - i\omega)k} \int_k^{+\infty} (e^x - e^k) p_S(T, x | \cdot) dx dk \\
&= \int_{-\infty}^{+\infty} p_S(T, x | \cdot) \left(e^x \int_{-\infty}^x e^{(\varepsilon - i\omega)k} dk - \int_{-\infty}^x e^{(\varepsilon - i\omega + 1)k} dk \right) dx \\
&= \left(\frac{1}{\varepsilon - i\omega} - \frac{1}{\varepsilon - i\omega + 1} \right) \int_{-\infty}^{+\infty} e^{(\varepsilon - i\omega + 1)x} p_S(T, x | \cdot) dx \\
&= \frac{1}{\varepsilon^2 + \varepsilon - i(2\omega\varepsilon + \omega) - \omega^2} \phi_S(T, \omega + (\varepsilon + 1)i, 0 | \cdot).
\end{aligned}$$

We can now invert the Fourier transform $\hat{c}_S(\omega)$ to obtain the subdiffusive call price

$$C_S(T, e^k, w, u, x_s, v_s) = \frac{e^{-\varepsilon k}}{\pi} \int_0^{+\infty} e^{i\omega k} \frac{\phi_S(T, \omega + (\varepsilon + 1)i, 0 | \cdot)}{\varepsilon^2 + \varepsilon - i(2\omega\varepsilon + \omega) - \omega^2} d\omega. \quad (3.7.19)$$

The Fourier-Laplace transform $\phi_S(T, \omega + (\varepsilon + 1)i, 0 | \cdot)$ at $z_2 = 0$ is evaluated using the approximation of Proposition 3.6 with the risk-neutral dynamics (3.5.9)-(3.5.10) (and hence with $-\eta\lambda_J m_J$ instead of μ and $\tilde{\kappa}, \tilde{\gamma}$ instead of κ, γ). The inversion of this transform in (3.7.19) is then performed efficiently using again a DFFT algorithm. Note that we implicitly assume here and in Section 3.7.1 on risk management that the truncation error of Proposition 3.6 is stable with respect to Fourier inversion. This should be rigorously verified in further research.

Contrary to risk management, pricing is done under the risk-neutral measure $\mathbb{Q}$ and we therefore need to find the corresponding parameters $\tilde{\kappa}$ and $\tilde{\gamma}$ under $\mathbb{Q}$ using the relation (3.5.11). Since there are no option data available in illiquid markets, there is no risk-neutral measure available from which we can estimate the market price of volatility risk θ_2 as well as the parameter η controlling the jump size distribution under $\mathbb{Q}$. Hence, we simply cannot derive the set of risk-neutral parameters based on available data. The best we can do is to use the observed log-return data under $\mathbb{P}$ and either assume a minimal change of measure by setting $\theta_2 = 0$ and $\eta = 1$ (leading to the same set of parameters) or obtain θ_2 and η based on standard values in the equity market (*e.g.* $\theta_2 = -1/\sigma$ as in Bollerslev et al. (2011)) or from a similar heavily-traded asset. For the sake of simplicity, we choose here the first option by considering the minimal change of measure $\theta_2 = 0, \eta = 1$. Moreover, since both of the above alternatives (3.7.17) and (3.7.19) for pricing European calls lead to the same values, we use the first method based on the conditional density $g(T, \tau | \cdot)$,

which is more direct in the case of the α -stable subordinator. We then compare for different maturities T the call price $C_S(T, K, 0, 0, x_0, v_0)$ under the SSVJ model with the call price $C(T, K, x_0, v_0)$ under the classical SVJ model (both at time $s = 0$). We also fix $K = e^{x_0} = 24.8$ such as to obtain the following ATM (at-the-money) term structure of option prices in Figure 3.7.2, which is also represented in terms of implied volatilities in Figure 3.7.3.

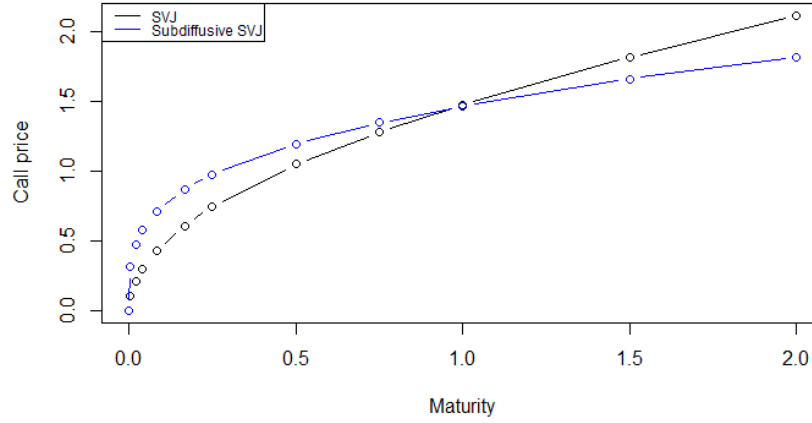


Figure 3.7.2: ATM European call prices on the ALQP stock under the classical SVJ model (black) and the SSVJ model (blue) for various maturities (with $K = 24.8$).

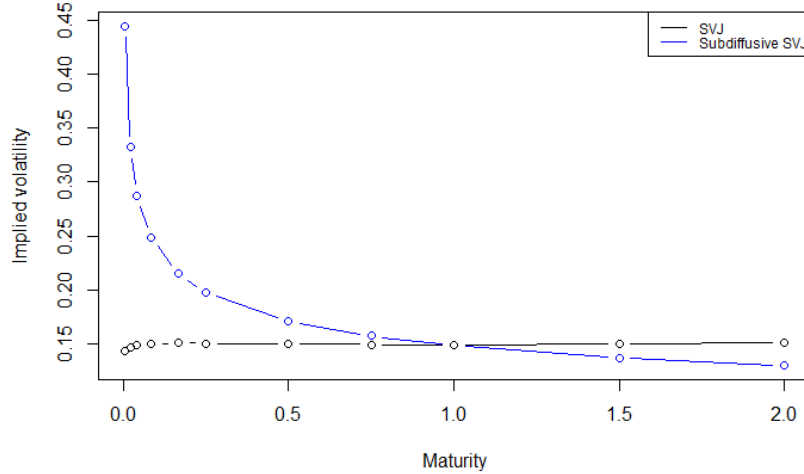


Figure 3.7.3: ATM term structure of implied volatilities of the ALQP stock price under the classical SVJ model (black) and the SSVJ model (blue) (with $K = 24.8$).

We clearly see an increased concavity of the curve of call prices under the subdiffusive model. Indeed, the lower the α , the more concave is this curve. Hence, considering the SSVJ model leads to higher call prices for small maturities (<1 year) and to lower prices for larger maturities (>1 year), compared to the classical SVJ model. Finally, we obtain a more realistic power-law shaped ATM term structure of implied volatilities when considering the SSVJ model (which is consistent with the empirical observations of Chapter 1, Section 1.5) while the SVJ model leads to a flatter ATM term structure. Hence, we see that the subdiffusive SVJ model deviates from the classical SVJ model in terms of option prices and implied volatilities, especially at short maturities.

3.8 Conclusion

This Chapter 3 proposes a new procedure for modeling illiquidity in financial markets. More precisely, we introduce the subdiffusive stochastic volatility jump model as a model of illiquid asset prices. It is built from the classical SVJ model that we time-change by an inverse subordinator S_t , controlling the periods of constant values in the asset price and the long-memory property of the log-returns. This SSVJ model allows to reproduce the motionless periods observed in the price of thinly-traded assets as well as the well-known volatility clustering and jump effects. The generated sample paths of this model are therefore clearly more in line with the observed prices in emerging markets and of small capitalization stocks, as depicted in Figure 3.1.1. As usual models of subdiffusions are developed in terms of FFPE, we show how to derive the FFPE satisfied by the conditional bivariate pdf of the time-changed log-stock price and its variance process, seen from time s . However, the Fourier-Laplace transform (3.4.19) of this pdf also admits a representation in terms of forward fractional PDE, which is far more easy to solve and on which we hence focus throughout the chapter. These fractional PDE are valid for a whole class of subordinators with no drift (encompassing *e.g.* the α -stable or Gamma subordinators) due to the fractional D-C derivative allowing for a lot of flexibility. Importantly, they can also be used at any subsequent time $s \geq 0$ thanks to our estimation procedure (and in particular the estimation of U_{S_s}) that allows to depart from the common Markovian valuation setup. We also show that this model admits a risk-neutral measure $\mathbb{Q}$, leading to an arbitrage-free and incomplete market. The SSVJ model is hence suitable for pricing various European-type financial products and leads to a more realistic term structure of option prices (and implied volatilities), compared with the classical SVJ model.

However, even if subdiffusions appear as good candidates for modeling illiquidity, options pricing in this framework remains a challenging task for two reasons. First, having no option quotes available in illiquid markets, we are forced to use the observed log-return data in order to estimate the parameters of our model.

We hence provide an econometric estimation of the model parameters based on the PMCMC procedure (with the Bayesian bootstrap filter). This method works well in our subdiffusive case and gives reliable results. Secondly, for option pricing and risk management, we have to solve numerically the forward fractional PDE (3.4.20) satisfied by the Fourier-Laplace transform ϕ_S . We use the series expansion of Proposition 3.6 since this approximation is the most adapted for solving our fractional PDE at $z_2 = 0$. A DFFT algorithm is finally applied to find the corresponding pdf associated with the obtained solution of the Fourier-Laplace transform ϕ_S and hence price options on subdiffusive assets. Note that throughout the practical part of the chapter, we have considered an α -stable process as subordinator and a compound Poisson process with Gaussian jump size for the jump process. However, since equation (3.4.20) is quite general, many other subordinators and (square integrable Lévy) jumps processes could be considered so as to fit the subdiffusive market as best as possible.

Finally, our estimation methodology allows to go beyond the simple setting in which valuation is restricted at the exact times of price moves. However, the SSVJ model requires the estimation of U_S which cannot be perfect at any time. Indeed, when the price is frozen at time t , the exact behavior of U_{S_t} cannot be obtained and therefore, its estimate is based on its average behavior conditionally on the estimation of S up to time t . Moreover, a second limitation of the SSVJ model concerns the application of our numerical scheme (Proposition 3.6) for option pricing and risk management. It is required for the truncation error of the generalized Taylor formula (3.7.12) to be stable with respect to Fourier inversion. As it goes beyond the scope of this chapter, this should be the subject of further research in computational mathematics. Finally, the zero risk-free rate and the minimal change of measure ($\theta_2 = 0$, $\eta = 1$) reduce to some extent the generality of our results.

3.A Proof of Proposition 3.4

1. Let us denote by $\mathcal{V}$ the subfiltration of $\mathbb{F}^A$ carrying exclusively information about the jump process and $\mathcal{G}_t := \mathcal{F}_0^{A \circ S} \vee \mathcal{F}_t^S \vee \mathcal{V}_{S_t}$. Using nested expectations, the mgf of $W_{2,S_t}^{\mathbb{Q}}$ under the equivalent measure $\mathbb{Q}$ is given by

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} \left[e^{\omega W_{2,S_t}^{\mathbb{Q}}} \middle| \mathcal{F}_0^{A \circ S} \right] &= \mathbb{E}^{\mathbb{P}} \left[Z_{S_t} e^{\omega W_{2,S_t}^{\mathbb{P}} + \omega \int_0^{S_t} \theta_{2,s} ds} \middle| \mathcal{F}_0^{A \circ S} \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[e^{\sum_{k=1}^{N_{S_t}^{\mathbb{P}}} \psi(Y_k^{\mathbb{P}}, \eta) - (\eta-1)\lambda_J S_t} \right. \\ &\quad \times \mathbb{E}^{\mathbb{P}} \left[e^{\omega W_{2,S_t}^{\mathbb{P}} + \omega \int_0^{S_t} \theta_{2,s} ds - \frac{1}{2} \int_0^{S_t} (\theta_{2,s}^2 + \theta_{\perp,s}^2) ds - \int_0^{S_t} \theta_{2,s} dW_{2,s}^{\mathbb{P}} - \int_0^{S_t} \theta_{\perp,s} dW_{\perp,s}^{\mathbb{P}}} \middle| \mathcal{G}_t \right] \middle| \mathcal{F}_0^{A \circ S} \right] \end{aligned}$$

Since the Brownian motion $W_2^{\mathbb{P}}$ is independent from the filtrations $\mathcal{V}$ and $\mathbb{F}^S$, we have that

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[e^{\omega W_{2,S_t}^{\mathbb{P}} + \omega \int_0^{S_t} \theta_{2,s} ds - \frac{1}{2} \int_0^{S_t} (\theta_{2,s}^2 + \theta_{\perp,s}^2) ds - \int_0^{S_t} \theta_{2,s} dW_{2,s}^{\mathbb{P}} - \int_0^{S_t} \theta_{\perp,s} dW_{\perp,s}^{\mathbb{P}}} \middle| \mathcal{G}_t \right] \\ = e^{\frac{1}{2}\omega^2 S_t} \mathbb{E}^{\mathbb{P}} \left[e^{-\frac{1}{2} \int_0^{S_t} ((\theta_{2,s} - \omega)^2 + \theta_{\perp,s}^2) ds - \int_0^{S_t} (\theta_{2,s} - \omega) dW_{2,s}^{\mathbb{P}} - \int_0^{S_t} \theta_{\perp,s} dW_{\perp,s}^{\mathbb{P}}} \middle| \mathcal{G}_t \right] \\ = e^{\frac{1}{2}\omega^2 S_t} \end{aligned}$$

To pass from the second to the last line, we use the fact that

$$e^{-\frac{1}{2} \int_0^{S_t} ((\theta_{2,s} - \omega)^2 + \theta_{\perp,s}^2) ds - \int_0^{S_t} (\theta_{2,s} - \omega) dW_{2,s}^{\mathbb{P}} - \int_0^{S_t} \theta_{\perp,s} dW_{\perp,s}^{\mathbb{P}}} \middle| \mathcal{F}_t^S$$

is a stochastic exponential and hence a true martingale thanks the chosen form (3.5.5) and (3.5.7) of the processes $\theta_{2,t}$ and $\theta_{\perp,t}$ with $\lambda \geq -\kappa/\sigma$ (see Theorem 3.5 of Wong and Heyde (2006) and its derivation). We recognize the mgf of a Brownian motion, time-changed by S_t . Since

$$e^{\sum_{k=1}^{N_{S_t}^{\mathbb{P}}} \psi(Y_k^{\mathbb{P}}, \eta) - (\eta-1)\lambda_J S_t} \middle| \mathcal{F}_t^S$$

is a martingale, nesting expectations leads to

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} \left[e^{\omega W_{2,S_t}^{\mathbb{Q}}} \middle| \mathcal{F}_0^{A \circ S} \right] &= \mathbb{E}^{\mathbb{P}} \left[e^{\frac{1}{2}\omega^2 S_t} \mathbb{E}^{\mathbb{P}} \left[e^{\sum_{k=1}^{N_{S_t}^{\mathbb{P}}} \psi(Y_k^{\mathbb{P}}, \eta) - (\eta-1)\lambda_J S_t} \middle| \mathcal{F}_0^{A \circ S} \vee \mathcal{F}_t^S \right] \middle| \mathcal{F}_0^{A \circ S} \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[e^{\frac{1}{2}\omega^2 S_t} \middle| \mathcal{F}_0^{A \circ S} \right] \end{aligned}$$

Therefore, we conclude that $W_{2,S_t}^{\mathbb{Q}} = W_{2,S_t}^{\mathbb{P}} + \int_0^{S_t} \theta_{2,s} ds$ is a time-changed Brownian motion under $\mathbb{Q}$. Using a similar argument, we also have that $W_{\perp,S_t}^{\mathbb{Q}} =$

$W_{\perp, S_t}^{\mathbb{P}} + \int_0^{S_t} \theta_{\perp, s} ds$ is a time-changed Brownian motion under $\mathbb{Q}$.

2. The second part of the proof concerning the jump process is identical to the one of Hainaut and Leonenko (2021), Proposition 5.1. $\square$

3.B Second simulated example

Based on the parameters (3.6.11), a second sample path of A_{S_t} (and hence of S_t , V_{S_t} , N_{S_t} and U_{S_t}) is simulated for $t \in [0, 5]$. The processes S and U_S are then estimated via our filtering technique based uniquely on the simulated values of A_S such as to confirm the accuracy of the estimation procedure as well as the conclusions of Section 3.6.1. We obtain the following figure 3.B.1.

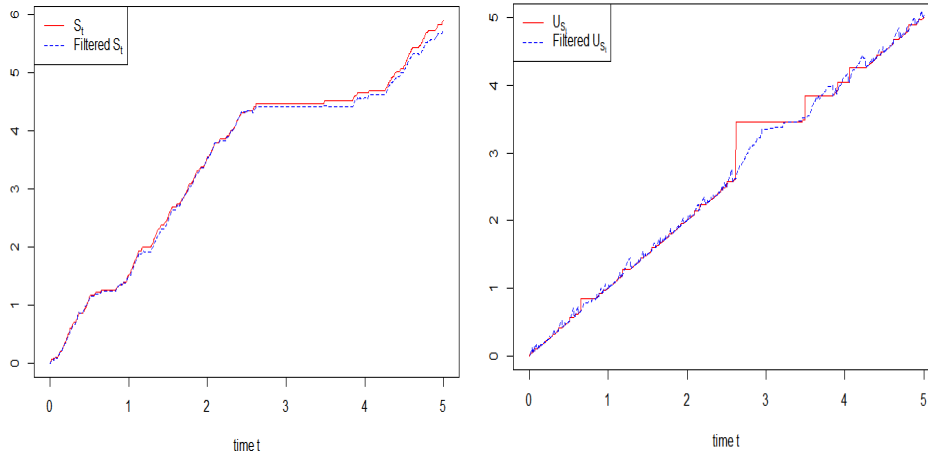


Figure 3.B.1: In red, second simulated sample path of the inverse subordinator S (left plot) and its related first-passage process U_S (right plot), compared with their filtered estimate in blue dotted line obtained via Algorithm 3.

3.C Existence of SSVJ moments

We here verify that the first four central moments of A_{S_T} under the SSVJ model are finite. First, based on equation 3.4.8, we can rewrite the unconditional moments

at maturity T as

$$\mathbb{E}^{\mathbb{P}}[A_{S_T}^k] = \int_0^\infty g(T, \tau) \mathbb{E}^{\mathbb{P}}[A_\tau^k] d\tau \quad (3.C.1)$$

where $\mathbb{E}^{\mathbb{P}}[A_\tau^k]$ is the unconditional central moment of order k of the classical SVJ model. Since all the moments of the jump process $(e^J - 1)N$ exist when J is Gaussian, it amounts to check for each k that the moments in the classical Heston case are finite for $\tau \geq 0$, as in Friz and Ressel (2010). We first define the moment explosion as

$$T_*(k) := \sup \{t \geq 0 : \mathbb{E}^{\mathbb{P}}[A_t^k] < \infty\},$$

and hence, $\mathbb{E}^{\mathbb{P}}[A_\tau^k]$ will be finite for $\tau < T_*(k)$ and infinite for $\tau \geq T_*(k)$. We say that no moment explosion occurs if $T_*(k) = \infty$.

- For $k = 1$, it is well known that $T_*(k) = \infty$ due the martingale property of $(A_t)_{t \geq 0}$ and hence that $\mathbb{E}^{\mathbb{P}}[A_\tau] < \infty$ for each $\tau \geq 0$. From equation (3.C.1) and since g is bounded, we finally have that $\mathbb{E}^{\mathbb{P}}[A_{S_T}]$ is finite for any $T \geq 0$.
- For $k = 2, 3, 4$, using Proposition 3.1 of Andersen and Piterbarg (2007) with the estimated set of parameters in Table 3.6.2, we find that $T_* = \infty$ (since $D \geq 0, a < 0$). Therefore, from (3.C.1), this again leads to the fact that $\mathbb{E}^{\mathbb{P}}[A_{S_T}^k]$ is finite for $k = 1, 2, 3$ and for any $T \geq 0$.

3.D Proof of Proposition 3.1 (iii).

By the definition (3.3.2) of S_t and the definition of the infimum, there is a random sequence $\tau_n \in \{\tau \geq 0 : U_\tau > t\}$ such that $\tau_n \downarrow S_t$. By right continuity of $(U_\tau)_{\tau \geq 0}$, we have $\lim_{n \rightarrow +\infty} U_{\tau_n} = U_{S_t}$. Since $U_{\tau_n} > t$ for all n , we also have $\lim_{n \rightarrow +\infty} U_{\tau_n} = U_{S_t} \geq t$. Hence, the case $x_1 \geq t$ is obvious since $U_{S_t} \geq t$ $\mathbb{P}$ -a.s.

Secondly, let $x_1 > t$ and $\omega \in \{S_{x_1} > S_t\}$. Then, there exists $S_\xi \in (S_t(\omega), S_{x_1}(\omega))$ such that $U_{S_\xi}(\omega) = \xi \leq x_1$. It follows that $U_{S_t(\omega)}(\omega) < U_{S_\xi}(\omega) = \xi \leq x_1$ and thus $\{U_{S_t} \geq x_1\} \cap \{S_{x_1} > S_t\} = \emptyset$. Since $(S_u)_{u \geq 0}$ is nondecreasing, we have

$$\begin{aligned} \{U_{S_t} \geq x_1\} &= \{U_{S_t} \geq x_1\} \cap \{S_{x_1} \geq S_t\} \\ &= (\{U_{S_t} \geq x_1\} \cap \{S_{x_1} > S_t\}) \cup (\{U_{S_t} \geq x_1\} \cap \{S_{x_1} = S_t\}) \\ &= \{U_{S_t} \geq x_1\} \cap \{S_{x_1} = S_t\} \end{aligned}$$

so that $\{U_{S_t} \geq x_1\} \subset \{S_{x_1} = S_t\}$. For the converse inclusion, let $\omega \in \{S_{x_1} = S_t\}$. Since $U_{S_t} \geq t$ $\mathbb{P}$ -a.s. from above, it implies that $x_1 \leq U_{S_{x_1}(\omega)}(\omega)$. Since $U_{S_{x_1}(\omega)}(\omega) = U_{S_t(\omega)}(\omega)$, we conclude that $\omega \in \{U_{S_t} \geq x_1\}$.

CHAPTER 4

A fractional Hawkes process for illiquidity modeling

In Chapter 4, we again consider a fractional process to directly model the illiquidity process associated with low-frequency observations of asset prices but this time, we shift our focus to the Amihud illiquidity measure. Indeed, the Amihud illiquidity measure has proved to be very popular in the economic and financial literature for measuring the illiquidity process of stocks and indices at daily time scales. None of the existing discrete-time illiquidity models in the literature are however adapted for reproducing peaks of illiquidity with long-memory, for risk management and for pricing liquidity-related derivatives. This Chapter 4 therefore proposes a new continuous-time paradigm for modeling illiquidity via a novel fractional Hawkes process in which the intensity process is ruled by a modified Mittag-Leffler excitation function. By considering a mean-reverting jump-diffusion model for the (log-)Amihud measure where jumps follow this modified fractional Hawkes process, we then manage to effectively reproduce the observed peaks of illiquidity in financial markets while introducing long-memory and tractability in the model. We can therefore use this model to perform risk management on illiquidity as well as to price illiquidity derivatives on the Amihud measure. In the continuity of Chapter 3, this chapter hence provides new tools for a better management of the illiquidity risk in financial markets.

Dupret, J. L., & Hainaut, D. A Fractional Hawkes
Process for Illiquidity Modeling. *Available at SSRN:*
<http://dx.doi.org/10.2139/ssrn.4162342>.

4.1 Introduction

Hawkes processes appear as good candidates for modeling the arrival of financial random events exhibiting a self-exciting behavior. They have been heavily used in finance for modeling *e.g.* the occurrence of trades at high frequency in Bacry and Muzy (2014), Bacry et al. (2015), Hardiman et al. (2013) and Cayé and Muhle-Karbe (2016); the firm defaults in a credit risk portfolio in Da Fonseca and Zaatour (2014) and Errais et al. (2010); or the jumps of asset prices in financial markets in Aït-Sahalia et al. (2015), Aït-Sahalia and Hurd (2016), Hawkes (2022) and Hainaut and Goutte (2019). The self-exciting behavior of Hawkes processes is characterized by the so-called excitation kernel, which is a decreasing function that provides the contribution of past jumps to the instantaneous probability of future jumps' occurrence. Several functions have been proposed in the literature as excitation kernel. The most common one is the exponential kernel, which enables a tractable and analytical treatment of Hawkes processes and hence provides explicit formulae for the moments, autocorrelation function and characteristic function of the number of jumps, see Aït-Sahalia et al. (2015), Da Fonseca and Zaatour (2014) or Errais et al. (2010). However, this exponential kernel cannot take into account the long-range dependence of past events on the number of jumps, which is typical in many financial applications such as in Bacry et al. (2015) for high-frequency trading, see also Hainaut (2022) for a review. Therefore, another widely-used excitation kernel proposed in the literature is the power-law one, as it allows for a slower decay compared to the exponential kernel and can thus better reproduce the observed long-term effects/memory in the number of jumps. This power-law kernel has also been extensively used in high-frequency modeling where the order flow is known to exhibit long-range dependence, as investigated in Bacry et al. (2012), Bacry et al. (2015), Hardiman et al. (2013). However, unlike the exponential kernel, the Hawkes process with power-law excitation function does not admit an analytical characteristic function of the number of jumps and thus lacks of tractability for many financial applications: valuation and risk management of portfolio credit derivatives as in Errais et al. (2010), option pricing for self-exciting jump diffusion models as in Hawkes (2022) or bond pricing under a self-exciting Hawkes dynamic for the short rate process as in Hainaut (2016). Hence, several authors have tried to fill this gap by introducing Hawkes models based on fractional dynamics. Hainaut (2020) proposes a fractional Hawkes process defined by a subdiffusive dynamic for the intensity process while Njike Leunga (2022) and Chen et al. (2021) consider respectively the Mittag-Leffler function and the derivative of the Mittag-Leffler function as excitation kernel. In the first part of this work, we hence review these fractional Hawkes processes and show that none of them are perfectly suited for the modeling of financial and economic events with long-range dependence. We then introduce a new fractional Hawkes process based on a modified Mittag-Leffler kernel that yields a tractable characteristic function while exhibiting long-memory.

We study its properties and show that this new fractional Hawkes process is the best suited for financial applications with long-range dependence, especially when the characteristic function of the number of jumps is required. Indeed, following Bäuerle and Desmettre (2020), we show how to rewrite our fractional Hawkes process with modified Mittag-Leffler kernel as a superposition of an infinite number of Markov processes. Discretizing the obtained representation, a semi-closed form solution is derived for the conditional transform of the number of jumps.

From the theoretical results developed in the first part of the chapter, we provide some practical applications of this modified fractional Hawkes process in the context of illiquidity modeling and show that it outperforms existing self-exciting models in terms of goodness of fit with observed illiquidity data in the market. Several econometric indicators have indeed been proposed for measuring the illiquidity process of stocks and indices using daily (or weekly) frequency data, see Goyenko et al. (2009) for a review. We here focus on the Amihud illiquidity measure as introduced in Amihud (2002), which has proved to be very popular in the empirical literature and to influence the cross-sectional asset returns through the so-called illiquidity premium, see the review of Amihud (2015). This premium arises due to investors taking into consideration illiquidity costs and pricing them into their expected returns so as to compensate for this market impact, see the general Introduction. Numerous econometric models in discrete time have then been proposed for modeling this Amihud measure, as in Amihud (2015), Brennan et al. (2013) or more recently Hafner et al. (2022). In this work, we shift of paradigm and consider a continuous-time process instead. More precisely, we model the (log-)Amihud measure as a mean-reverting jump diffusion where the jumps follow a fractional Hawkes process ruled by the modified Mittag-Leffler excitation function introduced in the first part of the chapter. This continuous-time approach helps us to obtain the characteristic function satisfied by the Amihud measure in quasi-closed form, leading to a time-efficient and accurate numerical computation of the moments, density and risk measures of the Amihud illiquidity measure. Therefore, this model enables time-efficient and exact risk management on illiquidity as well as to price financial instruments related to the Amihud measure¹. This way, we provide market participants with new tools to better manage and assess their risk of liquidity, to reduce their exposition to this risk and to express views on future liquidity conditions. Moreover, incorporating fractional Hawkes jumps in the model allows us to reproduce the observed peaks of illiquidity in financial markets and to show that these peaks effectively exhibit long-memory. Finally, by considering a multivariate setting for our fractional Hawkes process, we can study the contagion between the up and down shocks of this Amihud measure. This multivariate setting can also be adapted such as to simultaneously model the illiquidity process of multiple indices in financial markets as well as their contagion.

¹We refer to the original paper Dupret and Hainaut (2023a) for the introduction and pricing of liquidity-related derivatives, which are not included in this thesis for readability purposes.

The goal of Chapter 4 is therefore fourfold. We first introduce a new fractional Hawkes process with a modified Mittag-Leffler kernel and show that it exhibits better properties for the modeling of economic and financial events with long-range dependence, compared with the existing Hawkes models. We then propose a new paradigm based on a continuous-time process for studying the illiquidity of financial assets via the Amihud illiquidity measure. More precisely, working with a mean-reverting jump model for the log-Amihud measure with jumps following this modified fractional Hawkes process allows to easily reproduce the observed peaks of illiquidity in financial markets while introducing long-range dependence and tractability in the model. In a third step, we show via a Peaks Over Threshold procedure with likelihood maximization how to estimate the parameters of our mean-reverting fractional Hawkes model. By deriving the characteristic function satisfied by the Amihud measure in this model, we finally show in the last section of this chapter how to perform risk management on illiquidity by providing an efficient way to compute the distribution, moments and risk measures of the Amihud measure. We hence provide with this chapter new tools for a better understanding, management and assessment of the liquidity risk in financial markets.

4.2 Hawkes processes

The *Hawkes process* is a self-exciting point process introduced in Hawkes (1971). More precisely, it models a sequence of arrivals of some type over time such as earthquakes, gang violence, trade orders, price jumps or firm defaults. Each arrival excites the process in the sense that the chance of a subsequent arrival is increased for some time period after the initial event. As such, it is a non-Markov extension of the Poisson process.

We fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with right-continuous and complete filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, on which are defined a nonexplosive counting process N given by $N_t := \sum_{j \geq 1} \mathbb{1}_{\{\tau_j \leq t\}}$ and a point process L given by

$$L_t := \sum_{j \geq 1} O_j \mathbb{1}_{\{\tau_j \leq t\}}, \quad (4.2.1)$$

where the jump sizes $O_j \in \mathcal{F}_{\tau_j}$ are i.i.d. random variables and where $(\tau_j)_{j \geq 1}$ is a sequence of positive stopping times corresponding to the arrivals of jumps. These jumps represent market events such as peaks of illiquidity, defaults in a portfolio of firms or the occurrence of a trade. In a Hawkes setting, the processes N and L are directly specified through a conditional arrival rate or intensity λ as well as a distribution ν on $(0, \infty)$ for the jump size O (assumed to be strictly positive in this chapter). We suppose that the jump transform $\int_0^\infty e^{\xi o} d\nu(o)$ exists, is finite

for complex ξ and admits a finite derivative $\int_0^\infty o e^{\xi o} d\nu(o)$. The intensity then describes the conditional mean arrival rate of the Hawkes process and is assumed to follow a strictly positive stochastic process such that

$$\begin{aligned} \lambda_t \mid \mathcal{F}_t &= \lim_{h \rightarrow 0} \frac{\mathbb{E}[N_{t+h} \mid \mathcal{F}_t] - N_t}{h} = \lambda_0(t) + \int_0^t \eta f(t-u) dL_u \\ &= \lambda_0(t) + \eta \sum_{j: \tau_j < t} f(t - \tau_j) O_j \end{aligned} \quad (4.2.2)$$

where $\eta > 0$ and where $\lambda_0(t)$ is the positive background intensity function, describing the arrival of events triggered by a deterministic exogenous source. The function $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, called the self-excitation kernel of the process, is a decreasing function that provides the contribution to the intensity λ_t of a jump that occurred at a previous time $u < t$. Hence, each event increases the intensity by $\eta \lim_{h \rightarrow 0^+} f(h)O$, which then decays according to the excitation function $\phi := \eta f$ until the next event occurs to push it up again². This excitation kernel ϕ is decreasing so that more recent events have higher influence on the current intensity compared to events having occurred further away in the past. Note that in the self-excitation setting considered in this chapter, jumps of the Hawkes process always increase its intensity. Including negative impacts of jumps on the intensity λ_t is possible by replacing η in (4.2.2) with $\tilde{\eta}\lambda_u$, where $\tilde{\eta} \in \mathbb{R}$. This ensures to keep a non-negative intensity but at the cost of lessened tractability, see Da Fonseca and Malevergne (2021). From Bacry et al. (2015), we then have that the two-dimensional process $J = (L, N)^\top$ is a Hawkes process provided that the excitation kernel ϕ satisfies the following definition.

Definition 4.1. I. **Positive** : $\phi(t) \geq 0 \forall t$.

II. **Causal** : $\phi(t) = 0$ for $t < 0$.

III. **L^1 -integrable** : $\int_0^\infty \phi(s) ds < \infty$.

From equation (4.2.2), we clearly have that the higher the jump size O , the higher the effect of the jump on the intensity. Moreover, the intensity λ governs the common event time of N and L . However, while the jumps of the counting process N are unit-sized, the jumps of the point process L are drawn from the distribution ν . Finally, it is shown in Errais et al. (2010) that $N - \int_0^\cdot \lambda_s ds$ is a local martingale relative to $\mathbb{F}$ and $\mathbb{P}$.

As explained in Bacry et al. (2015), the L^1 norm of ϕ (denoted $\|\phi\|_1 := \int_0^\infty \phi(s) ds = \eta \|f\|_1$) can be interpreted as a branching ratio, *i.e.* the number of events generated by any parent event. Moreover, $\|\phi\|_1$ provides a direct measure of the fraction of

²This notation in terms of η and f will allow to emphasize the separate role from the clustering parameter η and the kernel f on the Hawkes process' properties, see Proposition 4.1.

endogenously triggered events within the whole population of events and is thus a measure of market reflexivity. Finally, even though a Hawkes process ruled by (4.2.2) is well-defined for any choice of kernel ϕ satisfying the three conditions of Definition 4.1, a stationary Hawkes process can be characterized as follows. The counting process N has asymptotically stationary increments and the intensity λ is asymptotically stationary if the kernel ϕ satisfies $\mathbb{E}(O)\|\phi\|_1 < 1$. In the multivariate extension of the Hawkes process that will be introduced in Section 4.3.4, the L^1 norm required for this stationary condition is replaced by the spectral radius, as explained in Bacry et al. (2015).

This setting is largely used in the financial literature to model the arrival of trade orders at high-frequency as the counting process N and the associated microstructure price as the loss process L (where O_j then denotes the order size of the j^{th} trade), see the review of Bacry et al. (2015). Similarly, firm defaults in a portfolio of credit risk can be modeled by the counting process N and the loss-given default by the random variable O , as in Errais et al. (2010) or Da Fonseca and Zaatour (2014). Finally, jumps in asset prices can also be studied using these Hawkes processes, see Aït-Sahalia et al. (2015), Hawkes (2022), Njike Leunga and Hainaut (2024). We now introduce a new fractional Hawkes process based on a modified Mittag-Leffler kernel. We will study this kernel throughout the chapter and compare it with standard choices of the excitation kernel f proposed in the literature.

4.2.1 The modified Mittag-Leffler fractional Hawkes process

This new fractional Hawkes process is based on the following excitation kernel, defined as the difference of a Mittag-Leffler (ML) function and its derivative, namely

$$f_\gamma(t) = E_\gamma(t^\gamma) - \frac{dE_\gamma(t^\gamma)}{dt} = E_\gamma(t^\gamma) - t^{\gamma-1}E_{\gamma,\gamma}(t^\gamma), \quad \gamma \in [1, 2], \quad t > 0, \quad (4.2.3)$$

where $E_\gamma(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(n\gamma+1)}$ is the one-parameter ML function and $E_{\gamma,\delta}(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(n\gamma+\delta)}$ the two-parameter ML function, both evaluated at $z = t^\gamma$. This kernel (4.2.3) with $\gamma \in [1, 2]$ is called the modified Mittag-Leffler (mML) kernel. From the expansion 18.1(22) in Bateman (1953), we can also easily derive the following asymptotics for our mML kernel (4.2.3) :

$$f_\gamma(t) \sim \begin{cases} 1 - \frac{t^{\gamma-1}}{\Gamma(\gamma)} \sim \exp\left[-\frac{t^{\gamma-1}}{\Gamma(\gamma)}\right], & t \rightarrow 0^+, \\ -\frac{t^{-\gamma}}{\Gamma(1-\gamma)}, & t \rightarrow \infty. \end{cases} \quad (4.2.4)$$

Therefore, we have that our newly introduced mML kernel interpolates between a stretched exponential for small times and a power-law with index $\gamma \in [1, 2]$ for large times. This kernel thus handles in an unified way the properties of the exponential

and power-law kernels (see Section 4.2.2 and Figure 4.2.1 below), providing both tractability and long-memory to the Hawkes process. This will hence lead to a better modeling and fine tuning of the short and long-term properties of the Hawkes process, as we will see in what follows. The intensity of the mML fractional Hawkes process is then defined in the following way

$$\lambda_t = \lambda_0 + \int_0^t f_\gamma(t-u) \alpha (\lambda_\infty - \lambda_u) du + \int_0^t f_\gamma(t-u) \eta dL_u, \quad (4.2.5)$$

where the background intensity function in (4.2.2) is chosen to be $\lambda_0(t) = \lambda_0 + \int_0^t f_\gamma(t-u) \alpha (\lambda_\infty - \lambda_u) du$, where $\lambda_0 > 0$ the intensity at initial time 0 and $\lambda_\infty > 0$ a fixed parameter controlling the long-run average intensity. Proposition 4.3 will show the existence and non-negativity of the intensity satisfying (4.2.5). The remainder of this section presents the most popular choices of decay function f discussed in the literature and compares them with the mML kernel (4.2.3). In particular, a numerical comparison is provided in Figure 4.2.1, which highlights the benefits of the mML kernel for the financial and economic modeling of events with long-term effects. The exact mathematical and statistical properties of the mML kernel (4.2.3) and of the mML Hawkes process (4.2.5) are then thoroughly studied in Section 4.3. In Section 4.4 and 4.5, we apply this new fractional Hawkes process for modeling (il)liquidity and we show that the short and long-term properties derived in Section 4.3 are perfectly suited compared with other excitation kernels for reproducing and studying the peaks of the (log-)Amihud illiquidity measure.

4.2.2 Standard Hawkes processes

We first review the two most frequent choices of decay function in the literature :

- **Exponential kernel :**

$$f_\alpha(t) = e^{-\alpha t}, \quad t \geq 0, \quad \alpha > 0, \quad (4.2.6)$$

where α is the rate at which the impact of an event decays over time and where the L^1 norm $\|\phi\|_1$ is equal to η/α . More precisely, a widely-used intensity process similar as our choice (4.2.5) with $\lambda_0, \lambda_\infty > 0$, is

$$\lambda_t = \lambda_\infty + e^{-\alpha t}(\lambda_0 - \lambda_\infty) + \eta \int_0^t e^{-\alpha(t-u)} dL_u \quad (4.2.7)$$

$$= \lambda_0 + \int_0^t f_\alpha(t-u) \alpha (\lambda_\infty - \lambda_0) du + \eta \int_0^t f_\alpha(t-u) dL_u, \quad (4.2.8)$$

which can be rewritten by differentiating the previous equation as

$$d\lambda_t = \alpha (\lambda_\infty - \lambda_t) dt + \eta dL_t. \quad (4.2.9)$$

As for the mML Hawkes intensity (4.2.5), the process λ does not blow up because the drift becomes negative whenever the intensity is above $\lambda_\infty > 0$ and prevents any explosion. The point process $J = (L, N)^\top$ is not Markov since it depends upon λ but the pair $X = (\lambda, J)^\top$ is well Markov in the state space $D = \mathbb{R}_+ \times (\mathbb{R}_+ \times \mathbb{N})$. Therefore, the infinitesimal generator of X , denoted $\mathcal{L}$, is the operator acting on a sufficiently regular function $f : D \rightarrow \mathbb{R}$ with continuous partial derivative $\partial f / \partial \lambda$ such that

$$\mathcal{L}f(x) = \lim_{h \rightarrow 0} \frac{\mathbb{E}_t[f(X_{t+h})] - f(x)}{h},$$

where $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t]$ and $X_t = x = (\lambda_t, (L_t, N_t))^\top$. For the Hawkes process (4.2.7), we can write

$$\begin{aligned} \mathcal{L}f(x) &= \alpha(\lambda_\infty - \lambda) \frac{\partial f}{\partial \lambda}(x) + \lambda \mathbb{E}[f(\lambda + \eta O, L + O, N + 1) - f(x)] \\ &= \alpha(\lambda_\infty - \lambda) \frac{\partial f}{\partial \lambda}(x) + \lambda \int_0^\infty (f(\lambda + \eta o, L + o, N + 1) - f(x)) \nu(\mathrm{d}o). \end{aligned}$$

Hence, Da Fonseca and Zaatour (2014) obtain the following Dynkin's formula for this Hawkes process with $h > 0$

$$\mathbb{E}_t[f(X_{t+h})] = f(X_t) + \mathbb{E}_t \left[\int_t^{t+h} \mathcal{L}f(X_u) \mathrm{d}u \right]. \quad (4.2.10)$$

Furthermore, since the process $X = (\lambda, J)^\top$ is a Markov process, we can express for $u = (u_\lambda, u_L, u_N) \in \mathbb{C}^3$ and $t \leq T$ its conditional transform $\mathbb{E}_t[e^{u \cdot X_T}]$ as $f(t, T, X_t)$, where f is a complex-valued function f on $[0, T] \times D$. Since f must satisfy the following PIDE

$$0 = \frac{\partial f}{\partial t}(t, T, x) + \mathcal{L}f(t, T, x), \quad (4.2.11)$$

with boundary condition $f(T, T, X_T) = e^{u \cdot X_T} = e^{u_\lambda \lambda_T + u_L L_T + u_N N_T}$, we have from Errais et al. (2010) that the point process $X = (\lambda, J)^\top = (\lambda, L, N)^\top$ is affine and that its conditional transform is given by

$$\mathbb{E}_t[e^{u \cdot X_T}] = \exp(a(t, T) + b(t, T)\lambda_t + u' \cdot J_t),$$

where $u' = (u_L, u_N) \in \mathbb{C}^2$ and the coefficient functions $a(t, T) := a(u', t, T)$ and $b(t, T) := b(u', t, T)$ satisfy the following Riccati ODEs

$$\begin{aligned} \frac{\partial b(t, T)}{\partial t} &= \alpha b(t, T) + 1 - \theta (\eta b(t, T) + u' \cdot (1, 0)^\top) e^{u' \cdot (0, 1)^\top}, \\ \frac{\partial a(t, T)}{\partial t} &= -\alpha \lambda_\infty b(t, T), \end{aligned}$$

with boundary conditions $a(T, T) = 0$, $b(T, T) = u_\lambda$ and where $\theta(\cdot)$ is the jump transform

$$\theta(\xi) = \int_0^\infty e^{\xi o} d\nu(o), \quad \xi \in \mathbb{C}.$$

This exponential kernel is used *e.g.* in Errais et al. (2010) in the context of portfolio credit risk, where the correlation of firm defaults is the main concern. Such specification of the Hawkes process allows the valuation of derivatives on debt portfolio in a highly tractable way thanks to the existence of closed-form formulae for its moments and characteristic function. However, in this framework of exponential excitation kernel, the autocovariance of the intensity decays exponentially with time. This feature is not adapted for modeling real phenomena that exhibit a long-term memory of past events such as trade occurrence, default events or illiquidity peaks (as we will see in Section 4.4). For example, it has been shown empirically in Hardiman et al. (2013) and in Bacry et al. (2012) that the exponential kernel is not the best suited for reproducing the occurrence of trades at high-frequency, suggesting a long memory nature of self-excitation phenomena at the microstructure level of price dynamics.

- Power-law kernel :

$$f_{\alpha, \gamma}(t) = \frac{\alpha}{(1 + \alpha t)^{1+\gamma}}, \quad \alpha > 0, \quad t \geq 0. \quad (4.2.12)$$

This kernel makes the pair (λ, J) non-Markov and we hence cannot rely on the above theory of infinitesimal generators to write the characteristic function of (λ, J) , see Bacry et al. (2015). In the case of portfolio credit derivatives, this prevents from valuating efficiently these contracts as done in Errais et al. (2010) with the exponential kernel (4.2.6) above. Similarly, this characteristic function is required for efficient risk management and option pricing¹ in the context of liquidity modeling, as we will see in Section 4.4-4.5. Other applications requiring this characteristic function include the pricing of options for self-exciting jump-diffusion models (see Aït-Sahalia et al. (2015)) or the pricing of bonds under Hawkes dynamics (see Hainaut (2016)). Nevertheless, the main benefit of this power-law kernel is to allow for a slower (power-law) decay than the exponential kernel so that it can be used for a better modeling of long-term effects/memory. We will indeed see in Figure 4.2.1 below that even for large maturities, there is a substantial contribution of past jumps in the intensity, inducing this lack of Markov property. This long-term power-law decay is *e.g.* observed at high-frequency in Hardiman et al. (2013) with $\gamma \approx 0$ (and thus a power-law exponent close to $1 + \gamma \approx 1$), which hence enables a better modeling of the long-range nature of offer and demand. We therefore want to build a kernel exhibiting such long-range dependence while having a characteristic function available in (semi-)closed form so that we can better manage the illiquidity risk and write options on liquidity-related instruments. Fractional Hawkes processes offer ways to achieve this, but first we show in the next section that none of the current fractional Hawkes in the literature fully satisfy the empirical properties observed in financial markets.

4.2.3 Fractional Hawkes processes

- Mittag-Leffler kernel :

$$f_\gamma(t) = E_\gamma(-t^\gamma), \quad t > 0, \gamma \in (0, 1]. \quad (4.2.13)$$

where $E_\gamma(z)$ is again the one-parameter ML function but this time with $\gamma \in (0, 1]$ and evaluated at $z = -t^\gamma$ (instead of $z = t^\gamma$ in (4.2.3)). For $t > 0$ and $\gamma \in (0, 1]$, the one-parameter ML function $E_\gamma(-t^\gamma)$ has the meaning of survival function for some positive random variable T with infinite mean, called the ML random variable. The choice of this kernel is motivated in a more general setting in Njike Leunga (2022) by the behavior of the ML function at short and long term. Indeed, it is commonly known that $E_\gamma(-t^\gamma)$ matches for $t \rightarrow 0$ with a stretched exponential with infinite negative derivative, whereas it matches as $t \rightarrow \infty$ with a negative power-law with index $\gamma \in (0, 1]$, see Mainardi (2020)

$$E_\gamma(-t^\gamma) \sim \begin{cases} \exp\left[-\frac{t^\gamma}{\Gamma(1+\gamma)}\right], & t \rightarrow 0^+, \\ \frac{t^{-\gamma}}{\Gamma(1-\gamma)} = \frac{\sin(\gamma\pi)}{\pi} \frac{\Gamma(\gamma)}{t^\gamma}, & t \rightarrow \infty. \end{cases} \quad (4.2.14)$$

As a consequence of the long-term power-law asymptotic, the process turns to be no longer Markov but of long-memory type. However, the L^1 norm is infinite which prevents from using this excitation function for modeling the arrival of financial random events. Indeed, using the probabilistic interpretation of the ML function as survival function for the positive ML random variable T , we find

$$\|f_\gamma(t)\|_1 = \int_0^\infty E_\gamma(-u^\gamma) du = \int_0^\infty \mathbb{P}(T \geq u) du = \mathbb{E}[T] = \infty$$

since from Theorem 3 (ii) in Lin (1998), the expected value of the the ML random variable T is infinite. In this case, each exogenous event generates infinitely many endogenous events and the average intensity explodes at large time scales, making the choice of the ML kernel not conceivable for modeling financial events. Indeed, this excitation function does not even define a Hawkes process since the third condition of Definition 4.1 is not satisfied.

- Mittag-Leffler's derivative kernel (dML) :

$$f_\gamma(t) = -\frac{dE_\gamma(-t^\gamma)}{dt} = t^{\gamma-1} E_{\gamma,\gamma}(-t^\gamma), \quad t > 0, \gamma \in (0, 1]. \quad (4.2.15)$$

where $E_{\gamma,\delta}(z)$ is the two-parameter ML function introduced in (4.2.3) but now evaluated at $z = -t^\gamma$. This kernel (4.2.15) is in fact the sign-changed first derivative of the ML kernel (4.2.13) and hence the pdf of the survival random variable T . It

has been introduced in Chen et al. (2021) as a non-Markov fractional Hawkes process with long-memory and with $\|f\|_1 = 1$, such that

$$t^{\gamma-1} E_{\gamma,\gamma}(-t^\gamma) \sim \begin{cases} +\infty & t \rightarrow 0^+, \\ \frac{-t^{-\gamma-1}}{\Gamma(-\gamma)} = \frac{\sin(\gamma\pi)}{\pi} \frac{\Gamma(\gamma+1)}{t^{\gamma+1}}, & t \rightarrow \infty. \end{cases} \quad (4.2.16)$$

Although the three conditions of Definition 4.1 are satisfied (making this fractional Hawkes process (4.2.15) well defined), we directly see in equation (4.2.15) with $\gamma < 1$ and in (4.2.16) that this kernel tends to infinity when $t \rightarrow 0^+$ and hence cannot capture the right behavior of the Hawkes process right after its jumps (given by $\eta \lim_{t \rightarrow 0^+} f_\gamma(t) O$). Moreover, the developments we will make for deriving the characteristic function of (λ, J) in Section 4.3.3 will not apply to this dML kernel (due to the divergent quantity (4.B.8)). As this characteristic function is required to perform risk management on illiquidity in Section 4.5, this prevents from using the ML's derivative kernel in our setting.

4.2.4 Comparison

We now compare numerically in Figure 4.2.1 the different choices of kernels above. We first observe at short times that the exponential, ML and mML kernels exhibit the same behavior. Indeed, both the ML and mML kernels behave as a stretched exponential for short times and we in fact have the same asymptotics for $t \rightarrow 0^+$ in (4.2.4) and (4.2.14) given that $\gamma \in [1, 2]$ for the mML kernel (4.2.3) whereas $\gamma \in (0, 1]$ in the ML kernel (4.2.13). We also clearly see the explosion of $f_\gamma(t)$ for $t \rightarrow 0^+$ in the dML kernel as already mentioned. At large times, we have instead that the power-law, dML and mML kernels exhibit the same decay. Indeed, the dML and mML kernels have the same long-term power-law decay since the asymptotics (4.2.4) and (4.2.16) for $t \rightarrow +\infty$ are equivalent (again from the condition $\gamma \in [1, 2]$ in the mML kernel and $\gamma \in (0, 1]$ in the dML kernel), which induces their non-Markov property. We also see that the ML kernel in dark blue decays as a power-law but much slower (at a rate γ instead of $1+\gamma$ for $\gamma \in (0, 1]$), compared with the other fractional kernels. This explains that the L^1 norm explodes for this process making it not conceivable for financial modeling. Finally, the exponential kernel decreases very rapidly to zero, preventing from using this kernel for modeling long-term effects since there is no contribution anymore of past jumps in the intensity for maturities $t > 1$. Therefore, thanks to its power-law behavior at large times and since it keeps tractable formulae (see next section) due to its fast decay at short time, our mML Hawkes process appears the best suited for modeling financial events with long-term dependence/memory. This will be confirmed empirically with Figure 4.A.2 in the context of the Amihud illiquidity measure. We now derive the exact mathematical and statistical properties of the mML kernel (4.2.3), required in Section 4.4 and 4.5 for modeling (il)liquidity with long-range dependence.

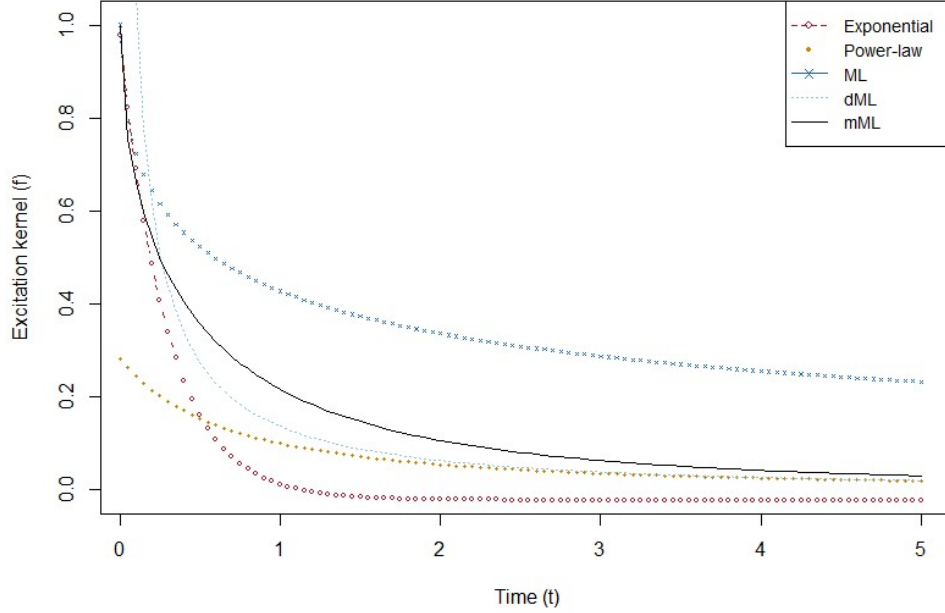


Figure 4.2.1: Comparison of the excitation kernels proposed in the literature with the mML kernel (4.2.3). In red and orange dotted lines, we have the exponential kernel with $\alpha = 3.12$ and the power-law kernel with $\alpha = 0.28, \gamma = 0.5$. In dark and light blue dotted lines, we have the ML and dML kernels with $\gamma = 0.5$. In black line, we have the mML kernel with $\gamma = 1.5$.

4.3 The modified ML fractional Hawkes process

4.3.1 Properties

We first recall that the mML kernel is defined in (4.2.3) as the difference of a ML kernel and its derivative, evaluated at t^γ with $\gamma \in [1, 2]$ and which admits the asymptotics (4.2.4) at short and large times. Moreover, we can directly expand the continuity of (4.2.3) in zero and hence obtain the right limit $\lim_{t \rightarrow 0^+} f_\gamma(t) = f_\gamma(0+) = 1$, ensuring a non-explosion and reasonable behavior of the Hawkes process at the occurrence of a jump. As depicted in Figure 4.2.1, subtracting the derivative of the ML kernel in the mML kernel (4.2.3) allows a faster rate of decay at large times, which is necessary to define properly a fractional kernel with long-range dependence and finite average (as we now show in Proposition 4.1 on the L^1 norm of $\phi_\gamma := \eta f_\gamma$). $\|\phi_\gamma\|_1$ is indeed equal to η and does not explode

anymore as in the classical ML kernel. This ensures that the mML Hawkes process is a suitable fractional Hawkes process for modeling financial and economic events with long-term effects.

Proposition 4.1. (i) *The L^1 norm of the modified Mittag-Leffler kernel (4.2.3) is given by*

$$\|f_\gamma\|_1 = \int_0^\infty f_\gamma(t) dt = \int_0^\infty (E_\gamma(t^\gamma) - t^{\gamma-1} E_{\gamma,\gamma}(t^\gamma)) dt = 1, \quad (4.3.1)$$

which thus leads to the L^1 norm $\|\phi_\gamma\|_1 = \eta \|f_\gamma\|_1 = \eta < \infty$.

(ii) *The mML fractional Hawkes process ruled by the excitation function (4.2.3) is well-defined according to Definition 4.1.*

The proof of this proposition is relegated in Appendix 4.B.1 and we now have that the parameter η controls the clustering property of the financial events modeled by this mML fractional Hawkes process. Secondly, the following Proposition 4.2 is crucial for deriving the conditional transform of (λ, J) under the mML fractional Hawkes process. This proposition will indeed be fundamental in the sequel of this chapter in order to perform risk management on illiquidity.

Proposition 4.2. *The modified Mittag-Leffler kernel (4.2.3) is completely monotone³ if and only if $\gamma \in [1, 2]$. This is equivalent to the existence of a representation of the kernel (4.2.3) in terms of a Laplace-Stieltjes integral with non-decreasing density and non-negative measure given by*

$$f_\gamma(t) = \int_0^\infty e^{-\omega t} \mu(d\omega), \quad (4.3.2)$$

where

$$\mu(d\omega) = \frac{-\sin(\gamma\pi) \omega^{\gamma-1} (1 + \omega)}{\pi (\omega^{2\gamma} - 2\omega^\gamma \cos(\gamma\pi) + 1)} d\omega. \quad (4.3.3)$$

We refer to Appendix 4.B.2 for the proof of this proposition.

4.3.2 Spectral representation

We first recall that the intensity of the mML Hawkes process is obtained by extending the exponential intensity model (4.2.8) with the mML kernel (4.2.3) in the following way

³A function $f : (0, +\infty) \rightarrow \mathbb{R}$ is completely monotone if it is C^∞ and if $(-1)^n f^{(n)}(x) \geq 0, \forall n \in \mathbb{N}$.

$$\lambda_t = \lambda_0 + \int_0^t f_\gamma(t-u) (\alpha(\lambda_\infty - \lambda_u) du + \eta dL_u). \quad (4.3.4)$$

This form for the intensity is in fact inspired by the affine Volterra structure of Abi Jaber et al. (2019) and Abi Jaber (2021) with Ornstein-Uhlenbeck (OU) drift coefficient, and justified by Proposition 4.3 below. Such mML kernel with affine Volterra structure for λ will lead in the sequel to a semi-closed form solution for the conditional transform of N and L . Moreover, from the spectral representation (4.3.2) and using Fubini's theorem, we can rewrite the mML intensity (4.3.4) as

$$\begin{aligned} \lambda_t &= \lambda_0 + \int_0^t \left(\int_0^\infty e^{-\omega(t-u)} \mu(d\omega) \right) (\alpha(\lambda_\infty - \lambda_u) du + \eta dN_u) \\ &= \lambda_0 + \int_0^\infty \left(\int_0^t e^{-\omega(t-u)} (\alpha(\lambda_\infty - \lambda_u) du + \eta dN_u) \right) \mu(d\omega) \\ &= \lambda_0 + \int_0^\infty Y_t^\omega \mu(d\omega). \end{aligned} \quad (4.3.5)$$

In the last line, we denote by Y_t^ω a Markov process satisfying for each index $\omega \in \mathbb{R}^+$ the SDE

$$dY_t^\omega = (\alpha(\lambda_\infty - \lambda_t) - \omega Y_t^\omega) dt + \eta dN_t, \quad Y_0^\omega = 0, \quad (4.3.6)$$

with solution

$$Y_t^\omega = \int_0^t e^{-\omega(t-u)} (\alpha(\lambda_\infty - \lambda_u) du + \eta dN_u). \quad (4.3.7)$$

From (4.3.5), we have that the intensity at time t is the sum of a background rate λ_0 and the expectation of arrival rates $(Y_t^\omega)_{\omega \in \mathbb{R}^+}$ with respect to a measure μ . We see this way that the mML fractional Hawkes intensity (4.3.4) can be written as a superposition of infinitely many Markov factors $(Y_t^\omega)_{\omega \in \mathbb{R}^+}$ mean reverting at different speeds ω towards the level $\alpha(\lambda_\infty - \lambda_t)/\omega$. Hence, the mML Hawkes process can be represented as an infinite-dimensional Markov process $(\lambda_t, (Y_t^\omega)_{\omega \in \mathbb{R}^+}, N_t)_{t \geq 0}$, which allows us to use the standard tools of stochastic calculus to derive its conditional transform from the affine property of equation (4.3.6).

Proposition 4.3. *There exists a unique non-negative (weak) solution for the intensity process λ satisfying (4.3.4).*

Proof. This directly comes from Theorem 2.13 of Abi Jaber (2021) since the kernel f_γ in (4.3.2) is completely monotone and taking $K = f_\gamma$, $g_0(t) = \lambda_0 + \int_0^t f_\gamma(t-u) \alpha \lambda_\infty du$ along with the semimartingale characteristics $b = 1 - \eta$, $c = 0$, $\nu = \nu$. See also Lemma 9 of Bondi et al. (2024) and Bondi et al. (2022). $\square$

4.3.3 Discretization and conditional transform

Following Bäuerle and Desmettre (2020), we now approximate the measure μ derived in (4.3.3) as a discrete measure with a finite number of atoms, allowing to move from an infinitely dimensional Markov process to a framework with finite dimensional Markovianity. For this purpose, we consider a partition $\mathcal{E}^{(n)} := \{0 < \xi_0^{(n)} < \xi_1^{(n)} < \dots < \xi_n^{(n)} < \infty\}$. The barycenter of μ on each interval $(\xi_k^{(n)}, \xi_{k+1}^{(n)})$ for $k = 0, \dots, n-1$ is given by

$$b_{k+1}^{(n)} = \left(\int_{\xi_k^{(n)}}^{\xi_{k+1}^{(n)}} \omega \mu(d\omega) \right) / \left(\int_{\xi_k^{(n)}}^{\xi_{k+1}^{(n)}} \mu(d\omega) \right). \quad (4.3.8)$$

For $k = 0, \dots, n-1$, the mass of any atom on each interval $(\xi_k^{(n)}, \xi_{k+1}^{(n)})$ is hence

$$m_{k+1}^{(n)} = \int_{\xi_k^{(n)}}^{\xi_{k+1}^{(n)}} \mu(d\omega) = \int_{\xi_k^{(n)}}^{\xi_{k+1}^{(n)}} \frac{-\sin(\gamma\pi) \omega^{\gamma-1} (1+\omega)}{\pi (\omega^{2\gamma} - 2\omega^\gamma \cos(\gamma\pi) + 1)} d\omega. \quad (4.3.9)$$

We cannot obtain analytical expressions for $m_{k+1}^{(n)}$ and $b_{k+1}^{(n)}$. However, the exact expressions for the barycenter and the mass of each interval are not useful in what follows. The discrete measure for a partition of size n is then defined as follows

$$\mu^{(n)}(\omega) := \sum_{k=1}^n m_k^{(n)} \delta_{b_k^{(n)}}(\omega), \quad (4.3.10)$$

where $\delta_{b_k^{(n)}}(\omega)$ is the Dirac measure located at point ω such that $\lim_{n \rightarrow \infty} \mu^{(n)}(\omega) = \mu(\omega)$. We consider that the following assumption holds for the partition $\mathcal{E}^{(n)}$:

- (i) $\xi_0^{(n)} \rightarrow 0$ and $\xi_n^{(n)} \rightarrow \infty$ when $n \rightarrow \infty$.
- (ii) $\max |\xi_{k+1}^{(n)} - \xi_k^{(n)}| \rightarrow 0$ when $n \rightarrow \infty$.
- (iii) $\mathcal{E}^{(n)} \subset \mathcal{E}^{(n+1)}$.

Considering the previous discretization setting, the following proposition connects the discretized measure $\mu^{(n)}$ to the true measure μ for any measurable function in the set $L^1(\mu)$ of integrable functions with respect to this measure μ .

Proposition 4.4. *For $n \in \mathbb{N}$, if a partition $\mathcal{E}^{(n)}$ satisfies (i)-(iii) and if a function $f : \mathbb{R} \rightarrow \mathbb{C}$ is in $L^1(\mu)$, then*

$$\lim_{n \rightarrow \infty} \int_0^\infty f(\omega) \mu^{(n)}(d\omega) = \int_0^\infty f(\omega) \mu(d\omega). \quad (4.3.11)$$

We refer to Appendix 4.B.3 for the proof, see also Bauerle and Desmettre (2020). We then fix $n \in \mathbb{N}$ and adopt the following notation when discretizing the mean-reverting factor Y^ω defined in (4.3.7) :

$$Y_t^{(k)} := Y_t^{b_k^{(n)}}, \forall k = 1, \dots, n,$$

where

$$dY_t^{(k)} = \left(\alpha \left(\lambda_\infty - \lambda_t^{(n)} \right) - b_k^{(n)} Y_t^{(k)} \right) + \eta O dN_t^{(n)},$$

and with $N^{(n)}$ a counting process of intensity $\lambda^{(n)}$. From equation (4.3.5), this discretized intensity $\lambda^{(n)}$ based on a partition of size n is given by

$$\lambda_t^{(n)} = \lambda_0 + \sum_{k=1}^n m_k^{(n)} Y_t^{(k)}. \quad (4.3.12)$$

From Proposition 4.4, we have that $\lim_{n \rightarrow \infty} \lambda_t^{(n)} = \lambda_t$ almost surely since we can rewrite equation (4.3.12) as⁴ $\lambda_t^{(n)} = \lambda_0 + \int_0^\infty Y_t^\omega \mu^{(n)}(d\omega)$ and since the function $\omega \in \mathbb{R}^+ \mapsto Y_t^\omega(w')$ is integrable with respect to the measure μ for fixed $w' \in \Omega$ as a sum of negative exponential functions, see equation (4.3.7). Finally, we obtain

$$d\lambda_t^{(n)} = \sum_{k=1}^n m_k^{(n)} dY_t^{(k)} = \sum_{k=1}^n m_k^{(n)} \left(\alpha \left(\lambda_\infty - \lambda_t^{(n)} \right) - b_k^{(n)} Y_t^{(k)} \right) + \eta O \sum_{k=1}^n m_k^{(n)} dN_t^{(n)}.$$

Now that we have discretized the mML fractional Hawkes process into a finite number of related Markov processes, we are able to obtain the main mathematical result of the chapter about its conditional transform with Proposition 4.5 and Corollary 4.1. These tools are indeed required when applying the mML Hawkes process for modeling illiquidity in Section 4.4 and 4.5.

Proposition 4.5. *The transform of $X_T^{(n)} = \left(\lambda_T^{(n)}, J_T^{(n)} \right)^\top = \left(\lambda_T^{(n)}, \left(L_T^{(n)}, N_T^{(n)} \right) \right)^\top$ conditionally to $\mathcal{F}_t$ at time $t \leq T$ with $u = (u_\lambda, u_L, u_N) \in \mathbb{C}^3$ is given by*

$$\begin{aligned} \mathbb{E}_t \left[e^{u \cdot X_T^{(n)}} \right] &= \mathbb{E}_t \left[e^{u_\lambda \lambda_T^{(n)} + u_L L_T^{(n)} + u_N N_T^{(n)}} \right] \\ &= \exp \left(a(t, T) + b(t, T) \lambda_t^{(n)} + u' \cdot {}^\top J_t^{(n)} + \sum_{k=1}^n c_k(t, T) Y_t^{(k)} \right), \end{aligned} \quad (4.3.13)$$

where $u' = (u_L, u_N) \in \mathbb{C}^2$ and where the coefficients $a(t, T) := a(u', t, T)$, $b(t, T) :=$

⁴Using $\int_0^\infty Y_t^\omega \mu^{(n)}(d\omega) = \sum_{k=1}^n m_k^{(n)} \int_0^\infty Y_t^\omega \delta_{b_k^{(n)}}(d\omega) = \sum_{k=1}^n m_k^{(n)} Y_t^{b_k^{(n)}} = \sum_{k=1}^n m_k^{(n)} Y_t^{(k)}$.

$b(u', t, T)$ and $c_k(t, T) := c_k(u', t, T)$ satisfy the following system of equations

$$\begin{aligned} a(t, T) &= \alpha \lambda_\infty \sum_{k=1}^n m_k^{(n)} \left(\int_t^T e^{-b_k^{(n)}(v-t)} b(v, T) dv \right), \\ \frac{\partial b(t, T)}{\partial t} &= \alpha \left(\sum_{k=1}^n m_k^{(n)} b(t, T) + \sum_{k=1}^n c_k(t, T) \right) \\ &\quad - \theta \left(b(t, T) \eta \sum_{k=1}^n m_k^{(n)} + u' \cdot (1, 0)^\top + \sum_{k=1}^n c_k(t, T) \eta \right) e^{u' \cdot (0, 1)^\top} + 1, \\ c_k(t, T) &= -b_k^{(n)} m_k^{(n)} \int_t^T e^{-b_k^{(n)}(v-t)} b(v, T) dv, \quad \forall k = 1, \dots, n. \end{aligned}$$

with boundary condition $b(T, T) = u_\lambda$ and where $\theta(\xi) = \int e^{\xi o} d\nu(o)$, $\xi \in \mathbb{C}$ is the transform of the jump distribution.

The proof of this proposition is given in Appendix 4.B.4. Passing to the limit $n \rightarrow \infty$ leads to the following corollary.

Corollary 4.1. *The transform of $X_T = (\lambda_T, J_T)^\top = (\lambda_T, (L_T, N_T))^\top$ conditionally to $\mathcal{F}_t$ at time $t \leq T$ with $u = (u_\lambda, u_L, u_N) \in \mathbb{C}^3$ is given by*

$$\begin{aligned} \mathbb{E}_t \left[e^{u \cdot X_T} \right] &= \mathbb{E}_t \left[e^{u_\lambda \lambda_T + u_L L_T + u_N N_T} \right] \\ &= \exp \left(\alpha \lambda_\infty \int_t^T f_\gamma(v-t) b(v, T) dv + b(t, T) \lambda_t \right. \\ &\quad \left. + u' \cdot J_t^\top - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} b(v, T) dv \right) \omega Y_t^\omega \mu(d\omega) \right), \end{aligned} \quad (4.3.14)$$

where $u' = (u_L, u_N) \in \mathbb{C}^2$ and where the coefficient function $b(t, T) := b(u', t, T)$ satisfies the following integro-differential equation

$$\begin{aligned} \frac{\partial b(t, T)}{\partial t} &= -\alpha \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(v, T) dv \\ &\quad - \theta \left(u' \cdot (1, 0)^\top - \eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(v, T) dv \right) e^{u' \cdot (0, 1)^\top} + 1, \end{aligned} \quad (4.3.15)$$

with boundary condition $b(T, T) = u_\lambda$.

The proof of this corollary is again relegated in Appendix 4.B.5. Based on the conditional transform (4.3.14), we can now obtain the moments of the process $X = (\lambda, J)^\top$ given in the following corollary.

Corollary 4.2. *The conditional expectation of X_T to $\mathcal{F}_t$ at time $t \leq T$ is given by*

$$\begin{aligned} \mathbb{E}_t[w \cdot X_T] = & \alpha \lambda_\infty \int_t^T f_\gamma(v-t) b'(0, v, T) dv + b'(0, t, T) \lambda_t + w' \cdot J_t^\top \\ & - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} b'(0, v, T) dv \right) \omega Y_t^\omega \mu(d\omega), \end{aligned}$$

for $w = (w_\lambda, w_L, w_N) \in \mathbb{R}^3$. We denote $w' = (w_L, w_N) \in \mathbb{R}^2$, $u' = (u_L, u_N) \in \mathbb{C}^2$ and $\chi = \mathbb{E}[O] = \int_0^\infty o \nu(do)$. The coefficient function

$$b'(0, t, T) := \left(w_L \frac{\partial}{\partial u_L} + w_N \frac{\partial}{\partial u_N} \right) b(u', t, T) \Big|_{u'=(0,0)} \quad (4.3.16)$$

satisfies

$$b'(0, t, T) = w_\lambda + (\chi\eta - \alpha) \int_t^T f_\gamma(v-t) b'(0, v, T) dv + (T-t) w' \cdot (\chi, 1)^\top.$$

See Appendix 4.B.6 for the proof. Similarly, higher order moments can be obtained by successive conditioning and differentiating. Note that Corollaries 4.1 and 4.2 drastically simplify at initial time $t = 0$ since $Y_0^\omega = 0$, $\forall \omega \in \mathbb{R}^+$. Finally, the ACF of the mML Hawkes process N can be obtained using the second order properties derived in Proposition 4 of Bacry et al. (2015) since the Laplace transform of f_γ is known in closed form via (4.B.3). This allows a comparison with the ACF of the number of jumps N based on the exponential and power-law kernels, see Figure 4.A.2.

4.3.4 A multivariate extension

We now want to extend the previous univariate setting by considering a multivariate mML fractional Hawkes process where the intensity $\lambda^{(i)}$ of each component process $i = 1, \dots, d$ may also depend on several of the other Hawkes components. For completeness, we also let each component's intensity $\lambda^{(i)}$ be influenced by exogenous risk factors in the form of a diffusive stochastic process. These intensities $\lambda^{(i)}$ can therefore replicate the diffusive fluctuations of the observed time series of market prices and market measures. We thus consider the $\mathbb{N}^d$ -valued counting process N and the $\mathbb{R}_+^d$ -valued point process L where the conditional intensity $\lambda_t^{(i)}$ of the i^{th} loss component $L_t^{(i)} := \sum_{j \geq 1} O_j^{(i)} \mathbb{1}_{\{\tau_j^{(i)} \leq t\}}$ is given by the following affine structure,

$$\lambda_t^{(i)} = \lambda_0^{(i)} + \int_0^t f_{\gamma^{(i)}}(t-u) \left(\alpha^{(i)} (\lambda_\infty^{(i)} - \lambda_u^{(i)}) du + \eta^{(i)} \cdot dL_u^\top + \sigma^{(i)} \sqrt{\lambda_u^{(i)}} dW_u^{(i)} \right) \quad (4.3.17)$$

where⁵ $\eta^{(i)} = (\eta^{(i,1)}, \eta^{(i,2)}, \dots, \eta^{(i,d)}) \in \mathbb{R}_+^d$ and where $W^{(i)}$ is a one-dimensional standard Brownian motion independent from the jump processes $L^{(i)}$, $\forall i = 1, \dots, d$, and from the other Brownian motions $W^{(j)}$, $\forall j \neq i$. The jump size $O^{(i)}$ of the i^{th} component $L^{(i)}$ follows a distribution $\nu^{(i)}$ on $(0, \infty)$. This specification enables the jump occurrence of one component process to have an impact on the intensities of other component processes, which facilitates the modeling of cross-excitation phenomena. For example, each component process $L^{(i)}$ can model the default process of a particular firm, where each default also has an impact on the default intensity of other firms. The setting (4.3.17) also allows the modeling of (il)liquidity peaks through the Amihud measure, as we will explore in the next section. Finally, this multivariate extension is required at high-frequency in order to model both the up and down mid-changes in prices, see Bacry and Muzy (2014) and Bacry et al. (2015). As in Section 4.3.3, we can rewrite the conditional intensity (4.3.17) as

$$\lambda_t^{(i)} = \lambda_0^{(i)} + \int_0^\infty Y_t^{\omega, (i)} \mu^{(i)}(d\omega), \quad (4.3.18)$$

for all $i = 1, \dots, d$ where $\mu^{(i)}(d\omega) = \frac{-\sin(\gamma^{(i)}\pi) \omega^{\gamma^{(i)}-1} (1+\omega)}{\pi(\omega^{2\gamma^{(i)}} - 2\omega^{\gamma^{(i)}} \cos(\gamma^{(i)}\pi) + 1)} d\omega$ and

$$dY_t^{\omega, (i)} = \left(-\omega Y_t^{\omega, (i)} + \alpha^{(i)} (\lambda_\infty^{(i)} - \lambda_t^{(i)}) \right) dt + \eta^{(i)} \cdot dL_t^\top + \sigma^{(i)} \sqrt{\lambda_t^{(i)}} dW_t^{(i)}, \quad (4.3.19)$$

with $Y_0^{\omega, (i)} = 0$. From the affine structure of equation (4.3.17) and from the representation (4.3.18) of each conditional intensity in terms of infinitely many mean-reverting factors $Y_t^{\omega, (i)}$, we can apply directly the same reasoning as in the univariate case (4.3.14) so as to obtain the conditional transform of the process $Z = (\lambda, J)^\top = (\lambda, (L, N))^\top$. The following Proposition 4.6 gives the exact expression of this conditional transform $\mathbb{E}_t[e^{u \cdot Z_T}]$ in the multivariate setting and is obtained as for Proposition 4.5 from the discretization scheme 4.3.3 based on the infinitesimal generator of Z . The conditional transform of the process $Z = (\lambda, J)^\top = (\lambda, (L, N))^\top$ is then defined on a state space $D \subseteq \mathbb{R}_+^d \times (\mathbb{R}_+^d \times \mathbb{N}^d)$ with $u_\lambda, u_L, u_N \in \mathbb{C}^d$ and $u = (u_\lambda, u_L, u_N) \in \mathbb{C}^{3d}$ and is given as follows.

Proposition 4.6. *The transform of $Z_T = (\lambda_T, (L_T, N_T))^\top$ conditionally to $\mathcal{F}_t$ at time $t \leq T$ with $u = (u_\lambda, u_L, u_N) \in \mathbb{C}^{3d}$ is given by*

$$\begin{aligned} \mathbb{E}_t[e^{u \cdot Z_T}] &= \mathbb{E}_t[e^{u_\lambda \cdot \lambda_T^\top + u_L \cdot L_T^\top + u_N \cdot N_T^\top}] = \exp \left[\beta(t, T) \cdot \lambda_t^\top + u_L \cdot L_t^\top + u_N \cdot N_t^\top \right. \\ &\quad \left. + \sum_{i=1}^d \left(\alpha^{(i)} \lambda_\infty^{(i)} \int_t^T f_{\gamma^{(i)}}(v-t) \beta^{(i)}(v, T) dv - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} \beta^{(i)}(v, T) dv \right) \omega Y_t^{\omega, (i)} \mu^{(i)}(d\omega) \right) \right] \end{aligned}$$

⁵Note that we can only select $m \leq d$ components by setting $\eta^{(i,j)} = 0$ for some j .

with the boundary condition $\beta(T, T) = u_\lambda$ and where each component $\beta^{(i)}(t, T)$, $i = 1, \dots, d$, of the vector of coefficient functions $\beta(t, T)$ satisfies the following integro-differential equation

$$\begin{aligned} \frac{\partial \beta^{(i)}(t, T)}{\partial t} = & -\alpha^{(i)} \frac{\partial}{\partial t} \int_t^T f_{\gamma^{(i)}}(v-t) \beta^{(i)}(v, T) dv - \frac{1}{2} \sigma^{(i)^2} \left(\frac{\partial}{\partial t} \int_t^T f_{\gamma^{(i)}}(v-t) \beta^{(i)}(v, T) dv \right)^2 \\ & - \theta^{(i)} \left(u_{L,i} - \sum_{j=1}^d \eta^{(j,i)} \frac{\partial}{\partial t} \int_t^T f_{\gamma^{(j)}}(v-t) \beta^{(j)}(v, T) dv \right) e^{u_{N,i}} + 1, \end{aligned} \quad (4.3.20)$$

where $\theta^{(i)}(\xi) = \int_0^\infty e^{\xi o} d\nu^{(i)}(o)$, $\xi \in \mathbb{C}$ is the transform of the loss size distribution of component i .

We refer to Appendix 4.B.7 for the proof. From the mathematical properties (and particularly the conditional transforms) derived in this section, we are now ready to apply the mML fractional Hawkes process for modeling the illiquidity associated with low-frequency observations of asset prices in financial markets.

4.4 Illiquidity modeling

4.4.1 Amihud measure

Liquidity is a fundamental property of a well-functioning market and lack of liquidity is generally at the heart of many financial crises and disasters. Common ways of measuring liquidity using market data include bid-ask spreads, effective spreads, realized spreads, depth and transaction volume. There is a big literature that uses such measures to compare market quality across time, sectors and before and after interventions of various sorts. Moreover, understanding and incorporating the behavior of liquidity in asset management and investment decision making is crucial for fund managers since their performance are strongly affected by the liquidity condition of their fund. It is indeed shown in Cao et al. (2013) that fund managers have the ability to time market liquidity and hence to adjust their portfolios' market exposure based on their insight of future liquidity conditions so as to bring substantially more return and performance to their fund. As proxy for illiquidity, we decide to focus on the Amihud illiquidity measure initially introduced in Amihud (2002). This measure has proved to be very popular in the empirical literature, easy to implement and by all accounts relatively robust. It has been shown to influence the cross-sectional asset returns through the so-called illiquidity premium, see *e.g.* Amihud (2015). This illiquidity premium indeed results from the fact that investors care about illiquidity costs and therefore price them in the expected returns for compensating this price/market impact.

Many econometric models in discrete time have been studied in the literature for

modeling illiquidity via the Amihud measure, see Amihud (2015), Brennan et al. (2013), Lou and Shu (2017) or more recently Hafner et al. (2022). We propose in this work a new approach by introducing instead a continuous-time process for modeling the log-Amihud measure. More precisely, this log-Amihud measure is modeled as a mean-reverting jump diffusion process with jumps following a fractional Hawkes process driven by a mML intensity with the multidimensional affine structure (4.3.17). This way, we want to show that peaks of illiquidity in financial markets effectively exhibit long-memory and that there is contagion between the up and down jumps of the log-Amihud measure. This model is also suitable for simultaneously studying the illiquidity process of multiple stocks and indices in financial markets. We will finally show in Section 4.5 that employing such a continuous-time process facilitates effective risk management on illiquidity.

The Amihud illiquidity measure A_t of a stock at time t is defined as

$$A_t = \frac{|R_t|}{V_t}, \quad (4.4.1)$$

where R_t is the stock return and V_t is the trading volume (in currency) at time t . We therefore see that the Amihud measure captures the fact that a stock is less liquid (more illiquid) if a given trading volume generates a larger move in its price. In practice, the measure is typically computed over low-frequency periods ranging from a day to a year by averaging the daily illiquidity ratio over the corresponding period p using $A_j := 1/p \sum_{k=0}^{p-1} |R_{t_j \cdot p - k}| / V_{t_j \cdot p - k}$ for $j = 1, \dots, N$. We consider for the remaining of this work a daily sampling period with $p = 1$ (hence $A_j = A_{t_j}$) and $\Delta t = t_j - t_{j-1} = 1/251$, $\forall j = 1, \dots, N$ with $t_N = T$ and $t_0 = 0$. Due to the very small order of magnitude of the Amihud measure ($\sim 10^{-16} - 10^{-12}$ for the most popular indices), it is customary to work with the logarithm of the Amihud measure instead as this standardizes its values, see Amihud (2002), Brennan et al. (2013), Lou and Shu (2017). We then denote $a_t := \log A_t$ in the sequel of the chapter. Some descriptive statistics of the stock returns, traded volumes, Amihud and log-Amihud measures are given for the FTSE 100 in the data section of Appendix 4.A, Table 4.A.1. Moreover, typical sample paths of the log-Amihud measure a_{t_j} and of $\Delta a_{t_j} = a_{t_j} - a_{t_{j-1}}$ ($j = 1, \dots, N$) are also pictured in Appendix 4.A for the FTSE 100 on Figure 4.A.1 with $N = 3266$ daily observations over a total period T of 13 years. These plots show that Δa_t , the log-Amihud measure in difference, randomly fluctuates around zero with some shocks that do not display any clear trend: negative abrupt movements followed by abrupt movements of either sign. We also observe some clustering with serial dependence, which is confirmed empirically with the autocorrelogram (ACF) of the log-Amihud measure a_t in Figure 4.A.2. As defined in Chapter 2.1 of Pipiras and Taqqu (2017), the sub-exponential decay of the ACF (and more precisely, its power-law decay at large lags) clearly indicates long-range dependence/memory in a_t . Moreover, we directly see that the ACF of the mML Hawkes process N is the only one able to reproduce both the exponential

decay at short lags and the power-law one at larger lags, as investigated in Section 4.2. This confirms the ability of this mML kernel to better reproduce the memory properties of the observed log-Amihud illiquidity process. Finally, the Histogram 4.A.3 of the FTSE100 log-Amihud increments Δa_t shows that they are close to Gaussian. However, standard Gaussian processes fail to reproduce the high kurtosis and ACF of the (log-)Amihud measure (see Table 4.A.1 and Fig. 4.A.2) and lead to unrealistic parameter values for reproducing the abrupt movements of Fig. 4.A.1

These observations justify to model the Amihud difference Δa_t as a mean-reverting Ornstein-Uhlenbeck (OU) process with two jump components (for up and down jumps) and to link the frequency of jumps to their size. Moreover, it is shown in Lou and Shu (2017) that it is judicious to model the log-Amihud measure in the same way as log-volatility, which is typically described by a mean-reverting OU process, see Andersen et al. (2001), Gatheral et al. (2018). Finally, by incorporating jumps with a mML intensity, the model can effectively reproduce both the observed illiquidity shocks as well as the long-range memory of the Amihud illiquidity measure, as highlighted above in the ACF 4.A.2 and in Lobato and Velasco (2000). We therefore consider the modified Mittag-Leffler Ornstein-Uhlenbeck (mML OU) process defined by the following dynamic of the log-Amihud measure $a_t = \log A_t$:

$$\begin{aligned} da_t &= \theta_1 (\theta_2 - a_t) dt + \theta_3 dW_t^{(a)} + dL_t^{(1)} - dL_t^{(2)} \\ &= \theta_1 (\theta_2 - a_t) dt + \theta_3 dW_t^{(a)} + O^{(1)} dN_t^{(1)} - O^{(2)} dN_t^{(2)}, \end{aligned} \quad (4.4.2)$$

where $L^{(1)}$ and $L^{(2)}$ are the loss processes defined by equation (4.2.1) of the up jump and down jump components, respectively. The conditional intensity $\lambda^{(i)}$ of the counting process $N^{(i)}$ is given by equation (4.3.17) and $W^{(a)}$ is a standard Brownian motion, independent from the two jump components and from $W^{(i)}$ ($i = 1, 2$). Recall that the variable $O^{(i)}$ representing the jump size/illiquidity shocks of each type is strictly positive with distribution $\nu^{(i)}$ on $(0, \infty)$. For simplicity, an exponential distribution is chosen for $\nu^{(i)}$ in the sequel. Note first that modeling the log-Amihud measure as a mML OU model guarantees the positivity of the Amihud measure and provides a semi-closed form formula for its conditional transform, as we will show in Section 4.4.3. Secondly, this model (4.4.2) allows to study the peaks of (il)liquidity of various assets while having a dependence and contagion between the up and down shocks of the log-Amihud measure. Finally, as already mentioned, our mML kernel (4.2.3) captures the long-range dependence in these (il)liquidity shocks while ensuring good properties of the intensity process.

4.4.2 Estimation procedure

We now aim at estimating the log-Amihud dynamic of the FTSE 100 index based on equation (4.4.2). Without loss of generality, we do not consider in equation (4.3.17) the Brownian component in the intensities $\lambda^{(1)}$ and $\lambda^{(2)}$ of the up and

down jump processes $N^{(1)}$ and $N^{(2)}$ (*i.e.* $\sigma^{(i)} = 0$). This then leads for $i = 1, 2$ to

$$\lambda_t^{(i)} = \lambda_0^{(i)} + \int_0^t f_{\gamma^{(i)}}(t-u) \left(\alpha^{(i)}(\lambda_\infty^{(i)} - \lambda_u^{(i)}) du + \eta^{(i,1)} dL_u^{(1)} + \eta^{(i,2)} dL_u^{(2)} \right) \quad (4.4.3)$$

For estimating this mML OU process, we use a modified Peaks Over Threshold (POT) procedure, as described in Hainaut and Moraux (2018). This method is divided into three steps. In the first step, from the time series of daily log-Amihud measures a_{t_j} ($j = 1, \dots, N$) of the FTSE 100 index (cfr Appendix 4.A), we estimate the parameters $(\theta_1, \theta_2, \theta_3)$ of a classical OU process via maximum likelihood without considering the jump components $L^{(1)}$ and $L^{(2)}$. The second step consists in detecting jumps given that jumps occur when the log-Amihud measure is above (for up jumps) or below (for down jumps) some thresholds. These two thresholds $\text{Tr}_j^{(1)}$ and $\text{Tr}_j^{(2)}$ are given for each observation $j = 1, \dots, N$ by

$$\text{Tr}_j^{(1)}(\alpha_1) := a_{t_j} + \theta_1 (\theta_2 - a_{t_j}) \Delta t + \theta_3 \phi^{-1}(\alpha_1) \sqrt{\Delta t}, \quad (4.4.4)$$

$$\text{Tr}_j^{(2)}(\alpha_2) := a_{t_j} + \theta_1 (\theta_2 - a_{t_j}) \Delta t + \theta_3 \phi^{-1}(\alpha_2) \sqrt{\Delta t}, \quad (4.4.5)$$

where $\phi^{-1}(\alpha_1)$ and $\phi^{-1}(\alpha_2)$ are the Normal quantiles at probability levels α_1 and α_2 . The levels α_1 and α_2 are chosen such that the sample without jumps $\left\{ a_{t_j} \mid a_{t_j} \in [\text{Tr}_j^{(1)}(\alpha_1), \text{Tr}_j^{(2)}(\alpha_2)] \right\}$ has a skewness and a kurtosis as closed as possible to the theoretical skewness and kurtosis of a OU process, which are respectively equal to 0 and 3 since the log-Amihud measure is Gaussian in this model, *i.e.*

$$\begin{aligned} \alpha_1^*, \alpha_2^* = \arg \min_{\alpha_1, \alpha_2} & \left(\text{Skew} \left\{ a_{t_j} \mid a_{t_j} \in [\text{Tr}_j^{(1)}(\alpha_1), \text{Tr}_j^{(2)}(\alpha_2)] \right\} \right)^2 \\ & + \left(\text{Kurt} \left\{ a_{t_j} \mid a_{t_j} \in [\text{Tr}_j^{(1)}(\alpha_1), \text{Tr}_j^{(2)}(\alpha_2)] \right\} - 3 \right)^2. \end{aligned}$$

The set of the $n_T^{(1)}$ up jumps and $n_T^{(2)}$ down jumps over $[0, T]$ is then given by

$$\left\{ \tau_1^{(1)}, \dots, \tau_{n_T^{(1)}}^{(1)} \right\} = \left\{ t_j \mid a_{t_j} > \text{Tr}_j^{(1)}(\alpha_1^*) \right\} \quad \text{and} \quad \left\{ \tau_1^{(2)}, \dots, \tau_{n_T^{(2)}}^{(2)} \right\} = \left\{ t_j \mid a_{t_j} < \text{Tr}_j^{(2)}(\alpha_2^*) \right\},$$

respectively. In the case of the FTSE 100, we find the following Figure 4.4.1 where we have in black the sample of log-Amihud differences Δa_t with jumps and in red the sample without jumps (with $\alpha_1^* = 0.9695$ and $\alpha_2^* = 0.0191$). Considering these thresholds, we re-estimate the parameter θ_3 such that the sample without jumps follows a OU process, *i.e.* $\Delta a_{t_j} = \hat{\theta}_1 (\hat{\theta}_2 - a_{t_j}) \Delta t + \theta_3 W_\Delta$ for all $a_{t_j} \in [\text{Tr}_j^{(1)}(\alpha_1^*), \text{Tr}_j^{(2)}(\alpha_2^*)]$ with $\hat{\theta}_1$ and $\hat{\theta}_2$ the estimators found at the first calibration step. When a jump is detected at time $\tau_l^{(i)}$ ($i = 1, 2$), the jump size $\Delta L_{\tau_l^{(i)}}^{(i)}$ is then equal to the absolute variation of the log-Amihud measure, *i.e.*

$$\Delta L_{\tau_l^{(i)}}^{(i)} = \left| \Delta a_{\tau_l^{(i)}} - \hat{\theta}_1 (\hat{\theta}_2 - a_{\tau_l^{(i)}-1}) \Delta t \right|.$$

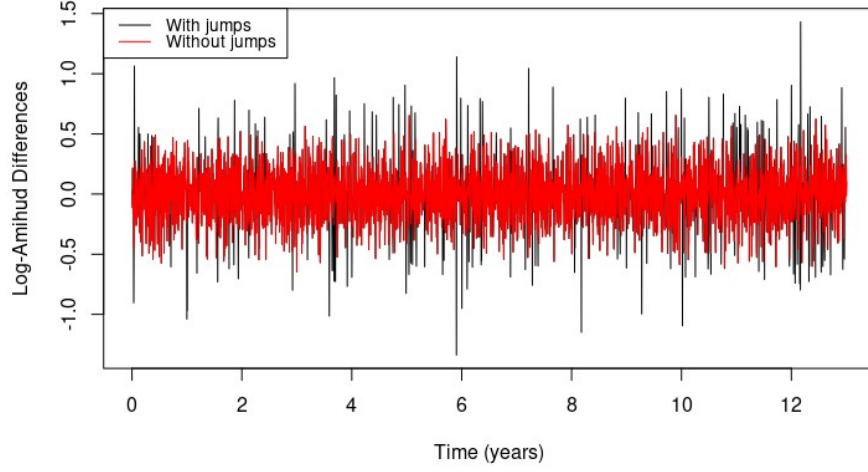


Figure 4.4.1: Daily log-Amihud measure in difference Δa_{t_j} ($j = 1, \dots, 3266$) for the FTSE 100 index over a period T of 13 years (cfr Appendix 4.A). In red, sample path without jumps and in black, sample path with jumps.

We also assume that the jump sizes $O^{(i)}$ ($i = 1, 2$) of the up and down jumps are independent and identically distributed exponential random variables with density $\nu^{(i)}$ on $\mathbb{R}^+$ given by

$$\nu^{(i)}(z | \rho^{(i)}) = \rho^{(i)} \exp(-\rho^{(i)} z), \quad z > 0,$$

with $\rho^{(i)} > 0$. The jump parameter $\rho^{(i)}$ is then estimated by log-likelihood maximization

$$\rho^{(i)} = \arg \max \sum_{l=1}^{n_T^{(i)}} \log \nu^{(i)} \left(\Delta L_{\tau_l^{(i)}}^{(i)} \mid \rho^{(i)} \right).$$

Based on the jump times $\{\tau_1^{(i)}, \dots, \tau_{n_T^{(i)}}^{(i)}\}$ of each type i , we can finally obtain via Embrechts et al. (2011) the log-likelihood of the mML Hawkes process with intensity (4.4.3), which is given by

$$\log L \left(\sum_{i=1}^d \sum_{l=1}^{n_T^{(i)}} \tau_l^{(i)} \mid \theta \right) = - \sum_{i=1}^d \int_0^T \lambda_s^{(i)} ds + \sum_{i=1}^d \sum_{l=1}^{n_T^{(i)}} \log \left(\lambda_{\tau_l^{(i)}}^{(i)} \right), \quad (4.4.6)$$

where $\theta = (\gamma^{(1)}, \alpha^{(1)}, \lambda_\infty^{(1)}, \lambda_0^{(1)}, \gamma^{(2)}, \alpha^{(2)}, \lambda_\infty^{(2)}, \lambda_0^{(2)}, \eta^{(1,1)}, \eta^{(1,2)}, \eta^{(2,1)}, \eta^{(2,2)})$. For the exponential kernel (4.2.6), the expression (4.4.6) can be drastically simplified as in Errais et al. (2010), which speeds up computation. In our mML case, knowing

the up and down jumps $\left\{\tau_1^{(i)}, \dots, \tau_{n_T^{(i)}}^{(i)}\right\}_{i=1,2}$ with their jump size $\Delta L_{\tau_l^{(i)}}^{(i)}$ allows us to use the spectral representation (4.3.18) and (4.3.19) along with their discretized counterparts (4.B.10) and (4.B.11) for $i = 1, 2$ based on a partition of size n (see Section 4.3.3). This leads to an efficient computation of the discretized intensity $\lambda_t^{(n,i)}$ and of the log-likelihood (4.4.6), based on

$$\lambda_{t_j}^{(n,i)} = \lambda_0^{(i)} + \sum_{k=1}^n m_k^{(n,i)} Y_{t_j}^{(k,i)}, \quad (4.4.7)$$

where $Y_0^{(k,i)} = 0$ and for $j = 0, \dots, N-1$:

$$\begin{aligned} Y_{t_{j+1}}^{(k,i)} = Y_{t_j}^{(k,i)} + \left(-b_k^{(n)} Y_{t_j}^{(k,i)} + \alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda_{t_j}^{(n,i)} \right) \right) \Delta t \\ + \eta^{(i,1)} \Delta L_{t_j}^{(n,1)} + \eta^{(i,2)} \Delta L_{t_j}^{(n,2)}. \end{aligned}$$

Restricting the number of factors n to 50 000 with $\Delta t = 0.001$ gives very accurate results and drastically reduces the computing time. A similar application of such technique with detailed convergence results can be found in Abi Jaber (2019). Note that since this discretized representation (4.4.7) cannot be written for the power-law and dML kernels (4.2.12)-(4.2.15), the computation of the log-likelihood (4.4.6) is much more slow in these cases. Finally, the estimated parameters of the mML intensity are then such that

$$\theta = \arg \max \log L \left(\sum_{i=1}^d \sum_{l=1}^{n_T^{(i)}} \tau_l^{(i)} \middle| \theta \right)$$

The full calibration procedure is summarized with the following Table 4.4.1. We first observe from representation (4.3.18)-(4.3.19) that the intensities $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$ mean revert towards a value $\hat{\lambda}_\infty^{(1)} = 4.721$ and $\hat{\lambda}_\infty^{(2)} = 3.386$, respectively (at a rate depending on $\hat{\alpha}^{(1)} = 6.048$ and $\hat{\alpha}^{(2)} = 0.451$). We also see that the parameters $\hat{\gamma}^{(1)}, \hat{\gamma}^{(2)}$ capturing the long-memory (*i.e.* the long-term power-law decay) of the illiquidity peaks in the log-Amihud measure are both very close to 1. Hence, this confirms the strong long-memory property of these illiquidity peaks. Moreover, based on equation (4.2.4) and Section 4.2.4, we have at large time scales the exact same power-law nature with exponent close to 1 for these illiquidity jumps as the one observed for mid-price changes at high frequency, which are based on the power-law kernel (4.2.12) with $\gamma \approx 0$ (and hence also with a power-law exponent close to $\gamma + 1 \approx 1$), cfr Bacry and Muzy (2014). This is also in line with the observed autocorrelogram 4.A.2 of a_t . Concerning the self-exciting parameters $\hat{\eta}^{(1,1)}$ and $\hat{\eta}^{(2,2)}$, we see that they are positive which confirms the presence of clustering effects in the intensity of the jumps. Moreover, these parameters are lower than the cross-excitation parameters $\hat{\eta}^{(1,2)}$ and $\hat{\eta}^{(2,1)}$ (which are also positive), leading to a predominant cross-excitation effect. Testing for stationarity in our bivariate setting (cfr Section 4.2), we obtain that the spectral radius $\|.\|$ of the matrix

Table 4.4.1: Comparison of the parameter estimates and standard errors of the mML OU model with the classical OU model for the log-Amihud measure. This table is obtained via the POT procedure on the FTSE 100 index based on a 13-year period of daily log-Amihud measures a_{t_j} ($j = 1, \dots, 3266$), see data section in Appendix 4.A.

	OU	Std. Error	mML OU	Std. Error
θ_1	27.789	2.187	27.789	2.187
θ_2	-34.290	1.701	-34.290	1.701
θ_3	4.196	0.054	4.313	0.055
Log-likelihood	-117.678		-197.832	
$\rho^{(1)}$			1.545	0.177
$\rho^{(2)}$			1.464	0.165
Log-likelihood			-91.85	
$\gamma^{(1)}$			1.029	0.064
$\lambda_0^{(1)}$			3.840	0.68
$\gamma^{(2)}$			1.024	0.021
$\lambda_0^{(2)}$			3.022	0.64
$\alpha^{(1)}$			6.048	0.264
$\alpha^{(2)}$			0.451	0.422
$\lambda_\infty^{(1)}$			4.721	0.812
$\lambda_\infty^{(2)}$			3.386	0.774
$\eta^{(1,1)}$			0.168	0.221
$\eta^{(1,2)}$			9.062	3.694
$\eta^{(2,1)}$			3.359	1.012
$\eta^{(2,2)}$			0.195	1.029
Log-likelihood			140.188	

$(\hat{\rho}^{(j)}\hat{\eta}^{(i,j)})_{i,j=1,2}$ is equal to 3.79. The stationary condition is therefore not satisfied based on our data. Finally, we see that we obtain high values of the speed of mean-reversion $\hat{\theta}_1$ and of the volatility $\hat{\theta}_3$. Computing the log-Amihud measure based on a weekly ($p = 5$) or monthly ($p = 21$) time window would lead to less extreme values. Another alternative frequently used in the literature consists in averaging this measure between different stocks/indices, as done in Amihud (2015).

The large standard errors of the self and cross-exciting parameters in Table 4.4.1 are explained by the limited number of up and down jumps used for estimating the model, leading to numerical instability. Finally, other distributions $\nu^{(i)}$ with positive support could have been chosen for the jump sizes $O^{(i)}$ (Log-normal, Gamma, etc). Determining the one leading to the best goodness of fit is let to further study.

We now consider other kernels as excitation function $f_{\gamma^{(i)}}(\cdot)$ in the jumps' intensity (4.4.3) and compare the results to Table 4.4.1. With the exponential kernel (4.2.6), we obtain the following estimated parameters $\hat{\alpha}^{(1)} = 75.103$, $\hat{\lambda}_0^{(1)} = 4.755$, $\hat{\lambda}_\infty^{(1)} = 4.055$, $\hat{\eta}^{(1,1)} = 0.051$, $\hat{\eta}^{(1,2)} = 19.901$, $\hat{\alpha}^{(2)} = 0.366$, $\hat{\lambda}_0^{(2)} = 2.619$, $\hat{\lambda}_\infty^{(2)} = 3.212$, $\hat{\eta}^{(2,1)} = 0.454$, $\hat{\eta}^{(2,2)} = 0.002$, leading to a log-likelihood (4.4.6) of 130.9. With the power-law kernel (4.2.12), we find the estimated parameters $\hat{\alpha}^{(1)} = 5.992$, $\hat{\gamma}^{(1)} = 12.528$, $\hat{\lambda}_0^{(1)} = 4.647$, $\hat{\eta}^{(1,1)} = 0.011$, $\hat{\eta}^{(1,2)} = 3.615$, $\hat{\alpha}^{(2)} = 0.161$, $\hat{\gamma}^{(2)} = 15.67$, $\hat{\lambda}_0^{(2)} = 2.750$, $\hat{\eta}^{(2,1)} = 2.467$, $\hat{\eta}^{(2,2)} = 0$ and a log-likelihood equal to 131.885. Using the AIC criterion to compare these models, we find that the mML Hawkes process gives the best results and hence better fits the time series of up and down jumps. Note that the ML and dML kernels also lead to a lower log-likelihood and a worse AIC criterion.

Finally, our multidimensional mML fractional Hawkes model also permits to simultaneous study the illiquidity process of multiple stocks and indices in financial markets as well as their contagion. For each index $i = 1, \dots, d$, the infinitesimal increments of the log-Amihud measure can be extended in the following way

$$da_t^{(i)} = \theta_1^{(i)} \left(\theta_2^{(i)} - a_t^{(i)} \right) dt + \theta_3^{(i)} dW_t^{(a,i)} + dL_t^{(i,up)} - dL_t^{(i,down)},$$

where $(W^{(a,i)})_{i=1,\dots,d}$ are independent Brownian motions (as well as independent from the jump components) and where the conditional intensity of the jump component $L^{(i,\ell)}$ of stock $i = 1, \dots, d$ and type $\ell = \{up, down\}$ is given by

$$\begin{aligned} \lambda_t^{(i,\ell)} = \lambda_0^{(i,\ell)} + \int_0^t f_{\gamma^{(i,\ell)}}(t-u) \left(\alpha^{(i,\ell)} (\lambda_\infty^{(i,\ell)} - \lambda_u^{(i,\ell)}) du \right. \\ \left. + \eta^{(i,up)} \cdot dL_u^{(up)} + \eta^{(i,down)} \cdot dL_u^{(down)} \right), \end{aligned}$$

with $\eta^{(i,\ell)} = (\eta^{(i,1,\ell)}, \dots, \eta^{(i,d,\ell)})$ and $L^{(\ell)} \in \mathbb{R}^d$. Therefore, in this model, the dependence between the log-Amihud measure of different stocks lies in the peaks of their illiquidity process and not from their diffusive factor which are considered as independent. A shock in the log-Amihud measure of one index has an impact on the intensity of the (up and down) jump component of the other indices. We can then apply the same POT procedure as above with the two thresholds (4.4.4) and (4.4.5) defined for each index i .

4.4.3 Characteristic function of the log-Amihud measure

Before performing risk management on the Amihud measure, we first need to derive the characteristic function $\varphi(t, T, \xi) := \mathbb{E} [\exp(\xi a_T) | \mathcal{F}_t]$ of the log-Amihud random variable a_T for $\xi \in \mathbb{C}$. We consider here the log-Amihud dynamic (4.4.2) with intensity (4.4.3), as proposed in the previous section where $i = 1, 2$ for the up and down jumps, respectively. We can then apply exactly the same proof as for Proposition 4.6 in order to obtain the following corollary.

Corollary 4.3. *The characteristic function of the log-Amihud measure a_T satisfying (4.4.2) conditionally to $\mathcal{F}_t$ at time $t \leq T$ with $\xi \in \mathbb{C}$ is given by*

$$\begin{aligned} \varphi(t, T, \xi) = \exp & \left[A(t, T) + B(t, T) \cdot {}^\top \lambda_t + D(t, T) a_t \right. \\ & \left. + \sum_{i=1}^2 \left(\alpha^{(i)} \lambda_\infty^{(i)} \int_t^T f_{\gamma^{(i)}}(v-t) B^{(i)}(v, T) dv - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} B^{(i)}(v, T) dv \right) \omega Y_t^{\omega, (i)} \mu^i(d\omega) \right) \right]. \end{aligned} \quad (4.4.8)$$

where the dynamic of $Y_t^{\omega, (i)}$ is given by equation (4.3.19) with $\sigma^{(i)} = 0$, $i = 1, 2$. The boundary conditions are given by $A(T, T) = 0$, $D(T, T) = \xi$ and $B(T, T) = (0, 0)$. Each component $B^{(i)}(t, T)$ of the coefficient vector $B(t, T)$ satisfies the following integro-differential equation for $i = 1, 2$:

$$\begin{aligned} \frac{\partial B^{(i)}(t, T)}{\partial t} = & -\alpha^{(i)} \frac{\partial}{\partial t} \int_t^T f_{\gamma^{(i)}}(v-t) B^{(i)}(v, T) dv \\ & - \theta^{(i)} \left(-\sum_{j=1}^2 \eta^{(j, i)} \frac{\partial}{\partial t} \int_t^T f_{\gamma^{(j)}}(v-t) B^{(j)}(v, T) dv + (-1)^{i-1} D(t, T) \right) + 1. \end{aligned} \quad (4.4.9)$$

The coefficient $D(t, T)$ is equal to

$$D(t, T) = \xi e^{\theta_1(t-T)},$$

Finally, the coefficient $A(t, T) = \int_t^T \left(\theta_1 \theta_2 D(v, T) + \frac{\theta_3^2}{2} D(v, T)^2 \right) dv$ is equal to

$$A(t, T) = \xi \theta_2 \left[1 - e^{\theta_1(t-T)} \right] + \frac{\theta_3^2 \xi^2}{4 \theta_1} \left[1 - e^{2\theta_1(t-T)} \right].$$

We do not give a dedicated proof for this corollary since it is directly derived from the proof of Proposition 4.6 in Appendix 4.B.7 and from the PIDE (4.2.11).

4.5 Risk management on illiquidity

We first use Corollary 4.3 with the obtained transform (4.4.8) to compute the moments of the Amihud measure A_T at time T for the FTSE 100 as well as several risk measures. Such measures can indeed help various financial institutions to better manage and assess their risk of liquidity. They can also help fund managers to determine which stock/index to buy as well as the optimal timing to do so. The different centered moments of the Amihud measure can be obtained easily thanks to the conditional transform $\varphi(t, T, k) = \mathbb{E}_t[e^{ka_T}] = \mathbb{E}_t[A_T^k]$ with $k \in \mathbb{N}$. The transform $\varphi(t, T, \xi)$ is obtained thanks to equation (4.4.8) of the previous corollary with the estimated parameters from Table 4.4.1.

Different risk measures for the Amihud measure A_T can also be computed via the probability density function (pdf) of a_T , which can be obtained efficiently via its Fourier representation. By setting $\xi = iz$, $z \in \mathbb{R}$ and inverting the characteristic function

$$\mathbb{E}_t[e^{iza_T}] = \int_0^\infty e^{iza_1} p(T, a_1 | t, a_2) da_1,$$

it is indeed possible to determine the conditional probability density $p(T, \cdot | t, a_2)$ of a_T , which is defined for $a_1, a_2 \in \mathbb{R}$ by

$$p(T, a_1 | t, a_2) := \frac{\partial}{\partial a_1} \mathbb{P}(a_T \leq a_1 | a_t = a_2).$$

We can then compute numerically the Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR). The VaR at level $\varepsilon \in (0, 1)$ is defined as in Chapter 2, eq. (2.7.2), by

$$\begin{aligned} VaR_\varepsilon(A_T) &= \inf \{q \in \mathbb{R} | \mathbb{P}(A_T \leq q) \geq \varepsilon\} \\ &= \inf \left\{ q \in \mathbb{R} \mid \int_{-\infty}^{\log q} p(T, a_1 | 0, a_2) da_1 \geq \varepsilon \right\}. \end{aligned}$$

We now use the following definition for the TVaR of the Amihud measure at level ε , $TVaR_\varepsilon(A_T) := \mathbb{E}[A_T | A_T > VaR_\varepsilon(A_T)]$, since we are interested in the right tail of the distribution. It can be rewritten in the following way

$$TVaR_\varepsilon(A_T) = \frac{1}{1 - \varepsilon} \int_{\log(VaR_\varepsilon(A_T))}^{+\infty} e^{a_1} p(T, a_1 | 0, a_2) da_1$$

The inversion of the Fourier transform is usually performed accurately and fast using a Discrete Fast Fourier Transform (DFFT) algorithm, see Appendix 2.F in Chapter 2. With the model parameters estimated from the POT procedure above in Table 4.4.1, we obtain in Table 4.5.1 the moments and risk measures of the Amihud measure A_T for $T = 1$ year. A comparison with the classical OU model

based again on the parameters $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ of Table 4.4.1 is also performed and the density of the Amihud and log-Amihud measures is depicted for each model in Figure 4.5.1 below.

Table 4.5.1: The first columns of the table give the first four central moments of the Amihud measure A_T at $T = 1$ under the classical OU model and the mML OU model with estimated parameters of Table 4.4.1, obtained from the daily log-Amihud measures of the FTSE 100 index (01/2009 - 01/2022), see Appendix 4.A. The two last columns give the respective $VaR_\varepsilon(A_T)$ and $TVaR_\varepsilon(A_T)$ at level $\varepsilon = 99.5\%$.

	Expectation	Variance	Skew.	Kurt.	VaR	TVaR
OU	1.52×10^{-15}	9.14×10^{-31}	2.14	12.12	5.69×10^{-15}	6.81×10^{-15}
mML OU	1.58×10^{-15}	1.82×10^{-30}	6.38	91.53	7.86×10^{-15}	1.24×10^{-14}

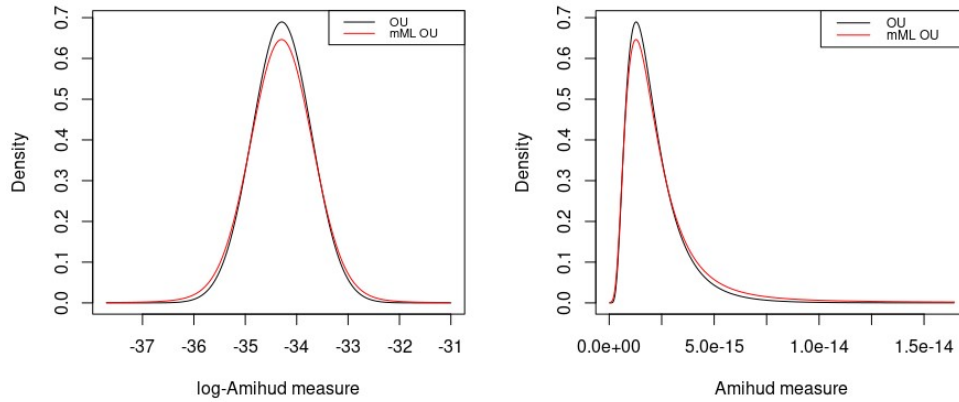


Figure 4.5.1: Density of the log-Amihud measure a_T (left) and of the Amihud measure A_T (right) for $T = 1$ year under the classical OU model (black line) and the mML OU model (red line), again based on the estimates of Table 4.4.1 for the FTSE 100 index.

We clearly see from Table 4.5.1 and Figure 4.5.1 that the mML OU model leads to a higher variability, a higher skewness and a higher kurtosis compared with the classical OU model. Since the intensity and size parameters θ and ρ between the up and down jumps are very close, the effect of the mML OU model is relatively small on the skewness (and mean) of the obtained distributions. However, the presence of these jumps leads to a strong effect on the kurtosis, the VaR and the TVaR, making this mML OU model more risky. Finally, if we compare Table 4.5.1 with the true descriptive statistics of the Amihud measure for the FTSE 100 index in Table 4.A.1 (third column), we see that the mML OU model indeed provides much

more accurate results than the OU model, confirming the ability of this model to better reproduce the observed illiquidity data.

4.6 Conclusion

This Chapter 4 proposes a new procedure for modeling illiquidity in financial markets. We indeed introduce a new fractional Hawkes process in which the intensity process is ruled by a modified Mittag-Leffler excitation function. The nice properties of the mML kernel in terms of L^1 norm and of short and long-term asymptotics make this fractional Hawkes process better suited for the modeling of financial/economic events with long-range dependence, compared with existing self-exciting models (exponential, power-law, ML or dML Hawkes processes). Moreover, thanks to its spectral representation which admits a useful discretization in terms of a finite number of Markov processes, the mML kernel offers sufficient tractability to our fractional Hawkes process so as to have a characteristic function available in semi-closed form. Finally, a multivariate extension of the mML fractional Hawkes process can also be derived easily so as to study the contagion between its components. Therefore, this new mML fractional setting is perfectly suited in the context of illiquidity modeling.

More precisely, we consider in this work the Amihud illiquidity measure, one of the most widely used proxy in illiquidity modeling, which is known to influence the asset returns through a liquidity premium that compensates for price impact. We hence provide a new paradigm for illiquidity modeling based on a mean-reverting (OU) jump process for the log-Amihud measure with jumps following this mML fractional Hawkes process. This so-called mML OU model simplifies many computations and leads to new tools for managing and reducing the illiquidity risk, while reproducing the observed peaks of illiquidity in financial markets and their long-memory property. In particular, having the characteristic function of this model in semi-closed form makes it also extremely convenient for risk management (and option pricing¹) on illiquidity so as to better assess and manage this risk. Finally, after having estimated our mML OU model via a POT procedure on FTSE 100's historical data, we show how this model leads to a more risky distribution of the Amihud measure compared with the classical OU model, which is more in accordance with empirical data.

In further research, we need to verify and generalize these results for a larger number of assets and indices while including a diffusive factor in their intensity process. We could also apply and test this new fractional Hawkes process in the context of credit risk portfolio, for option pricing in jump diffusion models or for the modeling of trade orders at high-frequency (and more generally for any financial/economic events with long-term effects).

4.A Data section

Table 4.A.1: Data statistics for the FTSE100's daily returns and trading volume between 01/2009 and 01/2022 (Yahoo Finance). The daily Amihud and log-Amihud measures are then computed from these data via equation (4.4.1) with $p = 1$.

	Stock return R_t	Volume V_t	Amihud measure A_t	Log-Amihud measure a_t
Mean	8.98×10^{-5}	8.31×10^8	1.55×10^{-15}	-34.66
Variance	1.16×10^{-4}	8.28×10^{16}	2.95×10^{-30}	1.44
Skewness	-0.74	1.85	5.85	-0.91
Kurtosis	9.31	7.62	86.25	1.37
Range	0.21	3.75×10^9	3.72×10^{-14}	10.13

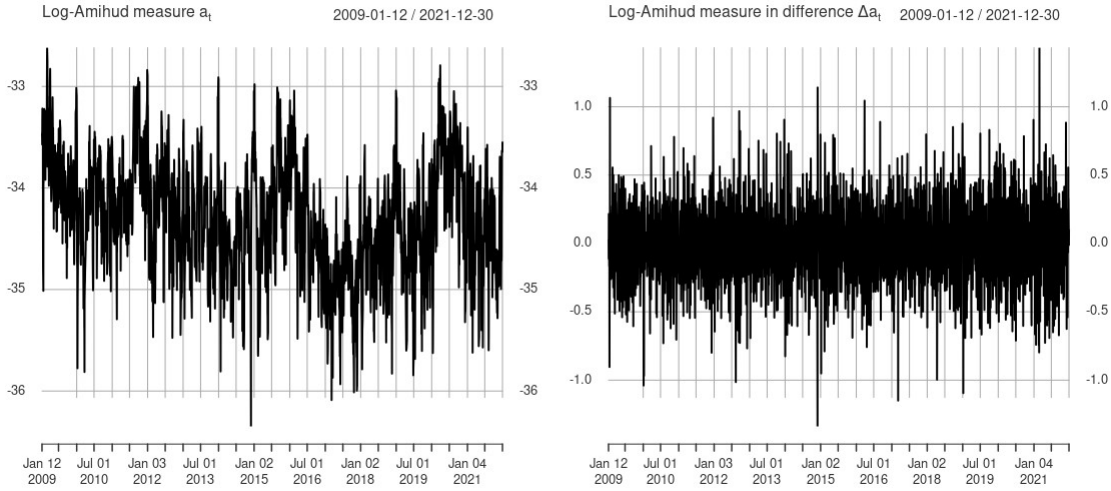


Figure 4.A.1: Daily log-Amihud measure in level a_{t_j} (left) and in difference Δa_{t_j} (right) with $j = 1, \dots, 3266$ for the FTSE 100 index over a period T of 13 years (01/2009 - 01/2022).

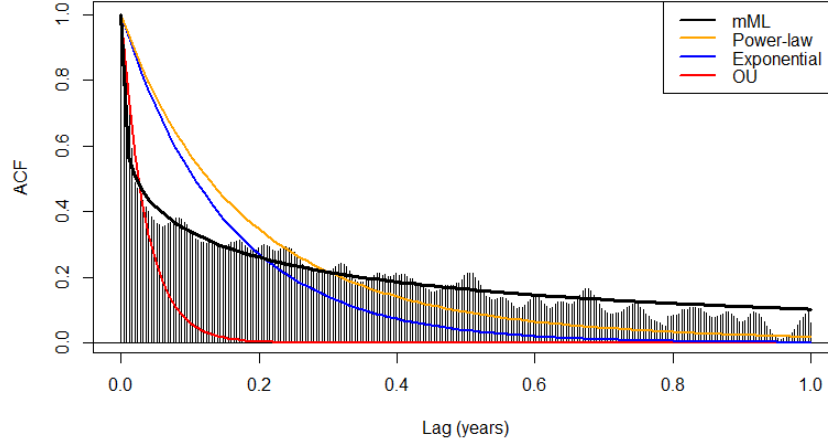


Figure 4.A.2: Autocorrelogram (ACF) of the daily log-Amihud measure a_{t_j} ($j = 1, \dots, 3266$) for the FTSE 100 based on a 13-year period (2009–2022) with lag up to 1 year. In red, ACF of the classical OU model (without jumps) with the estimated parameters of Table 4.4.1. The three remaining curves depict the ACF of the Hawkes process N using the second order properties from Proposition 4 of Bacry et al. (2015) with parameters chosen such as to best fit the observed ACF: in black, ACF of N with the mML kernel and $\gamma = 1.07$; in blue, ACF of N with the exponential kernel and $\alpha = 6.53$; in orange, ACF of N with the power-law kernel and $\alpha = 1$, $\gamma = 5.58$.

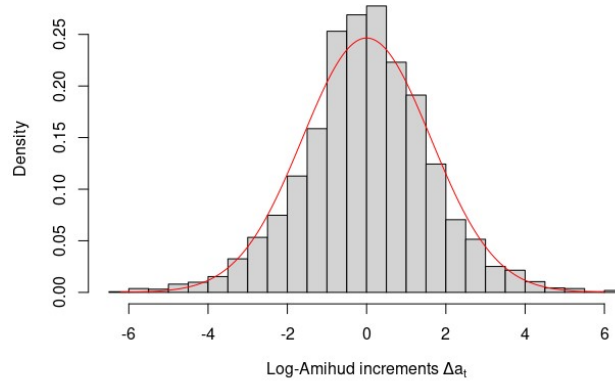


Figure 4.A.3: Histogram of the daily log-Amihud increments Δa_{t_j} ($j = 1, \dots, 3266$) for the FTSE 100 index based on a 13-year period (01/2009 - 01/2022) with the Gaussian fit on these data superimposed in red.

4.B Proofs

4.B.1 Proof of Proposition 4.1.

(i) In order to derive the L^1 norm of the mML kernel (4.2.3), we first recall some probabilistic results derived from Kyprianou (2006) (page 216 and Theorem 6.16). Let us first define an α -stable process without negative jumps with parameter $\alpha \in (1, 2]$ denoted U and an α -stable process without positive jumps $\tilde{U} := -U$. We can now define the supremum $\bar{U}_t := \sup\{U_s, s \leq t\} = -\inf\{\tilde{U}_s, s \leq t\}$ and $\mathbf{e}_1$ an exponential random variable with parameter 1, independent from U . Therefore, we have from Kyprianou (2006) that the Wiener-Hopf factorization leads for $\lambda \geq 0$ to

$$\mathbb{E}[e^{-\lambda \bar{U}_{\mathbf{e}_1}}] = \frac{\lambda - 1}{\lambda^\alpha - 1}, \quad (4.B.1)$$

and hence, integrating by parts, to

$$\int_0^\infty e^{-\lambda t} \mathbb{P}(\bar{U}_{\mathbf{e}_1} \geq t) dt = \frac{\lambda^{\alpha-1}}{\lambda^\alpha - 1} - \frac{1}{\lambda^\alpha - 1}.$$

Moreover, it is well-known from Mainardi (2020) that the function $E_\gamma(t^\gamma)$ has the following explicit Laplace transform

$$\mathcal{L}[E_\gamma(t^\gamma)](\lambda) = \int_0^\infty e^{-\lambda t} E_\gamma(t^\gamma) dt = \frac{\lambda^{\gamma-1}}{\lambda^\gamma - 1},$$

and integrating by parts, it comes for the modified Mittag-Leffler kernel (4.2.3) :

$$\begin{aligned} \mathcal{L}[f_\gamma(t)](\lambda) &= \int_0^\infty e^{-\lambda t} (E_\gamma(t^\gamma) - t^{\gamma-1} E_{\gamma,\gamma}(t^\gamma)) dt = \frac{\lambda^{\gamma-1}}{\lambda^\gamma - 1} - \frac{1}{\lambda^\gamma - 1} \\ &= \int_0^\infty e^{-\lambda t} \mathbb{P}(\bar{U}_{\mathbf{e}_1} \geq t) dt, \end{aligned} \quad (4.B.2)$$

when identifying γ and α . Hence, we finally find by inverting the Laplace transform (4.B.2) :

$$\mathbb{P}(\bar{U}_{\mathbf{e}_1} \geq t) = E_\gamma(t^\gamma) - t^{\gamma-1} E_{\gamma,\gamma}(t^\gamma).$$

We therefore find the following L^1 norm for the modified Mittag-Leffler kernel

$$\|f_\gamma\|_1 = \int_0^\infty \mathbb{P}(\bar{U}_{\mathbf{e}_1} \geq t) dt = \mathbb{E}[\bar{U}_{\mathbf{e}_1}] = 1.$$

The expectation is obtained by differentiating the moment generating function (mgf) (4.B.1) with respect to λ and evaluating the derivative at $\lambda = 0$. This thus

leads to $\|\phi_\gamma\|_1 = \eta \|f_\gamma\|_1 = \eta < \infty$.

(ii) Combined with the fact that ϕ_γ is also positive and causal, the three conditions of Definition 4.1 are verified and we finally have that the mML fractional Hawkes process ruled by the excitation function (4.2.3) is well-defined. $\square$

4.B.2 Proof of Proposition 4.2.

The proof of Proposition 4.2 is direct from the Titchmarsh formula for the inversion of Laplace transforms, see Gross (1950). We first have that the Laplace transform of the kernel (4.2.3) is given by equation (4.B.2) :

$$\mathcal{L}_{f_\gamma}(\lambda) = \int_0^\infty e^{-\lambda t} f_\gamma(t) dt = \frac{\lambda^{\gamma-1} - 1}{\lambda^\gamma - 1}. \quad (4.B.3)$$

We have that this function is analytic such that $\mathcal{L}_{f_\gamma}(\lambda) \rightarrow 0$ as $|\lambda| \rightarrow \infty$ and $\mathcal{L}_{f_\gamma}(\lambda) = o(|\lambda|^{-1})$ as $|\lambda| \rightarrow 0$ uniformly in every sector $|\arg \lambda| < \pi$, and that for $r > 0$ and $\theta \in (-\pi, \pi)$:

$$\Re(e^{i\theta/2} \mathcal{L}_{f_\gamma}(re^{i\theta}) + e^{-i\theta/2} \mathcal{L}_{f_\gamma}(re^{-i\theta})) = 2 \frac{\cos(\theta/2)(\lambda^{2\gamma-1} + 1) - \cos(\theta\alpha)(\lambda^{\gamma-1} + \lambda^\gamma)}{\lambda^{2\gamma} - 2\lambda^\gamma \cos(\theta\gamma) + 1} \geq 0,$$

with $\alpha = \gamma - 1/2 \in [1/2, 3/2]$. Therefore, the assumptions of Hirschman and Widder (2012) are satisfied and we then have the following "*restoring pair property*". Under these assumptions, the knowledge of $\mathcal{L}_{f_\gamma}(\lambda)$ enables one to find the inverse transform $\mu(\omega)$ of $f_\gamma(t)$, that is

$$f_\gamma(t) = \int_0^\infty e^{-\omega t} \mu(\omega) d\omega. \quad (4.B.4)$$

Following Gross (1950), we have from (4.B.3) and (4.B.4) that

$$\mathcal{L}_{f_\gamma}(\lambda) = \int_0^\infty \frac{\mu(\omega) d\omega}{\omega + \lambda}. \quad (4.B.5)$$

This is a Stieltjes equation, which is inverted by

$$\begin{aligned} \mu(\omega) &= \frac{1}{\pi} \operatorname{Im} \left[\mathcal{L}_{f_\gamma}(\omega e^{-i\pi}) \right] = \frac{\omega^{2\gamma-1} \sin(-\pi(\gamma-1) + \pi\gamma) - \omega^{\gamma-1} \sin(-\pi\gamma + \pi) + \omega^\gamma \sin(-\pi\gamma)}{\pi(\omega^{2\gamma} - 2\omega^\gamma \cos(\gamma\pi) + 1)} \\ &= \frac{-\sin(\gamma\pi) \omega^{\gamma-1} (1 + \omega)}{\pi(\omega^{2\gamma} - 2\omega^\gamma \cos(\gamma\pi) + 1)}. \end{aligned} \quad (4.B.6)$$

$\square$

4.B.3 Proof of Proposition 4.4

Using (4.3.10) in the first line and (4.3.9) to pass to the second line, along with assumptions (i) – (iii) on the partition in Section 4.3.3, we find

$$\begin{aligned} \int_0^\infty f(\omega) \mu^{(n)}(d\omega) &= \sum_{k=1}^n m_k^{(n)} \int_0^\infty f(\omega) \delta_{b_k^{(n)}}(d\omega) = \sum_{k=1}^n m_k^{(n)} f(b_k^{(n)}) \\ &= \sum_{k=0}^{n-1} \int_{\xi_k^{(n)}}^{\xi_{k+1}^{(n)}} f(b_{k+1}^{(n)}) \mu(d\omega) \\ &= \int_{\xi_0^{(n)}}^{\xi_n^{(n)}} f(t^{(n)}(\omega)) \mu(d\omega), \end{aligned}$$

where $t^{(n)}(\omega) = \sum_{k=0}^{n-1} b_{k+1}^{(n)} \mathbb{1}_{[\xi_k^{(n)}, \xi_{k+1}^{(n)}]}(\omega)$. Given that f is continuous and that $\lim_{n \rightarrow \infty} t^{(n)}(\omega) = \omega$ under assumptions (i) – (iii), passing to the limit leads to the announced result. $\square$

4.B.4 Proof of Proposition 4.5.

We first express the conditional transform $\mathbb{E}_t \left[e^{u \cdot X_T^{(n)}} \right] = \mathbb{E}_t \left[e^{u_\lambda \lambda_T^{(n)} + u_L L_T^{(n)} + u_N N_T^{(n)}} \right]$ as $f(t, T, X_t^{(n)}, (Y_t^{(k)})_{k=1, \dots, n})$. Moreover, the infinitesimal generator of $X^{(n)}$ is given by

$$\begin{aligned} \mathcal{L}f(x) &= \frac{\partial f}{\partial \lambda^{(n)}}(x) \sum_{k=1}^n m_k^{(n)} \left(\alpha \left(\lambda_\infty - \lambda^{(n)} \right) - b_k^{(n)} Y^{(k)} \right) + \sum_{k=1}^n \frac{\partial f}{\partial Y^{(k)}}(x) \left(\alpha \left(\lambda_\infty - \lambda^{(n)} \right) - b_k^{(n)} Y^{(k)} \right) \\ &\quad + \lambda^{(n)} \int_0^\infty \left(f \left(\lambda^{(n)} + \eta o \sum_{k=1}^n m_k^{(n)}, L^{(n)} + o, N^{(n)} + 1, (Y^{(k)} + \eta o)_{k=1, \dots, n} \right) - f(x) \right) \nu(d\omega), \end{aligned} \tag{4.B.7}$$

where $X_t^{(n)} = x = (\lambda_t^{(n)}, (L_t^{(n)}, N_t^{(n)}))^\top$. Since f must again satisfy the PIDE (4.2.11) with boundary condition $f(T, T, X_T^{(n)}) = e^{u \cdot X_T^{(n)}} = e^{u_\lambda \lambda_T^{(n)} + u_L L_T^{(n)} + u_N N_T^{(n)}}$, we have that the conditional transform of the point process $X_t^{(n)}$ is given by the exponential function

$$f \left(t, T, X_t^{(n)}, (Y_t^{(k)})_{k=1, \dots, n} \right) = \exp \left(a(t, T) + b(t, T) \lambda_t^{(n)} + u' \cdot {}^\top J_t^{(n)} + \sum_{k=1}^n c_k(t, T) Y_t^{(k)} \right),$$

where $u' = (u_L, u_N)$. The coefficient functions $a(t, T) := a(u', t, T)$ and $b(t, T) := b(u', t, T)$ are time dependent functions with boundary conditions $a(T, T) = 0$ and

$b(T, T) = u_\lambda$. The function $c_k(t, T) := c_k(u', t, T)$ is a function that depends again on time but also on the position $k = 1, \dots, n$, within the partition $\mathcal{E}^{(n)}$, and which satisfies the boundary condition $c_k(T, T) = 0, \forall k = 1, \dots, n$. Differentiating f with respect to $t, \lambda^{(n)}$ and $Y^{(k)}$ leads to

$$\begin{aligned}\frac{\partial f}{\partial t} &= f \left(\frac{\partial a(t, T)}{\partial t} + \frac{\partial b(t, T)}{\partial t} \lambda^{(n)} + \sum_{k=1}^n \frac{\partial c_k(t, T)}{\partial t} Y^{(k)} \right), \\ \frac{\partial f}{\partial \lambda^{(n)}} &= f b(t, T), \\ \frac{\partial f}{\partial Y^{(k)}} &= f c_k(t, T), \quad \forall k = 1, \dots, n,\end{aligned}$$

and the variation of f due to the occurrence of a jump of size o is given by

$$f \left(e^{b(t, T) \eta \circ \sum_{k=1}^n m_k^{(n)} + u' \cdot (1, 0)^\top o + u' \cdot (0, 1)^\top + \sum_{k=1}^n c_k(t, T) \eta \circ} - 1 \right).$$

Injecting these equations in (4.B.7) and in the PIDE (4.2.11), we have

$$\begin{aligned}0 &= \frac{\partial a(t, T)}{\partial t} + \frac{\partial b(t, T)}{\partial t} \lambda^{(n)} + b(t, T) \sum_{k=1}^n m_k^{(n)} \left(\alpha \left(\lambda_\infty - \lambda^{(n)} \right) - b_k^{(n)} Y^{(k)} \right) \\ &\quad + \sum_{k=1}^n c_k(t, T) \left(\alpha \left(\lambda_\infty - \lambda^{(n)} \right) - b_k^{(n)} Y^{(k)} \right) + \sum_{k=1}^n \frac{\partial c_k(t, T)}{\partial t} Y^{(k)} \\ &\quad + \lambda^{(n)} \left[\left(\int e^{b(t, T) \eta \circ \sum_{k=1}^n m_k^{(n)} + u' \cdot (1, 0)^\top o + \sum_{k=1}^n c_k(t, T) \eta \circ} \nu(\mathrm{d}o) \right) e^{u' \cdot (0, 1)^\top} - 1 \right].\end{aligned}$$

From which we find the following system of PDE

$$\begin{aligned}\frac{\partial a(t, T)}{\partial t} &= -b(t, T) \sum_{k=1}^n m_k^{(n)} \alpha \lambda_\infty - \sum_{k=1}^n c_k(t, T) \alpha \lambda_\infty, \\ \frac{\partial b(t, T)}{\partial t} &= b(t, T) \sum_{k=1}^n m_k^{(n)} \alpha + \sum_{k=1}^n c_k(t, T) \alpha \\ &\quad - \theta \left(b(t, T) \eta \sum_{k=1}^n m_k^{(n)} + u' \cdot (1, 0)^\top + \sum_{k=1}^n c_k(t, T) \eta \right) e^{u' \cdot (0, 1)^\top} + 1, \\ \frac{\partial c_k(t, T)}{\partial t} &= b(t, T) m_k^{(n)} b_k^{(n)} + c_k(t, T) b_k^{(n)}, \quad \forall k = 1, \dots, n,\end{aligned}$$

where $\theta(\cdot)$ is again the transform of the jump distribution.

The previous system of equation finally gives with $a(T, T) = 0$ and $c_k(T, T) = 0$:

$$\begin{aligned} a(t, T) &= \alpha \lambda_\infty \int_t^T \left(\sum_{k=1}^n m_k^{(n)} b(v, T) + \sum_{k=1}^n c_k(v, T) \right) dv, \\ \frac{\partial b(t, T)}{\partial t} &= \alpha \left(\sum_{k=1}^n m_k^{(n)} b(t, T) + \sum_{k=1}^n c_k(t, T) \right) \\ &\quad - \theta \left(b(t, T) \eta \sum_{k=1}^n m_k^{(n)} + u' \cdot (1, 0)^\top + \sum_{k=1}^n c_k(t, T) \eta \right) e^{u' \cdot (0, 1)^\top} + 1, \\ c_k(t, T) &= -b_k^{(n)} m_k^{(n)} \int_t^T e^{-b_k^{(n)}(v-t)} b(v, T) dv, \quad \forall k = 1, \dots, n, \end{aligned}$$

and $a(t, T)$ can be rewritten

$$\begin{aligned} a(t, T) &= \alpha \lambda_\infty \sum_{k=1}^n m_k^{(n)} \left(\int_t^T b(v, T) dv - b_k^{(n)} \int_t^T \int_v^T e^{-b_k^{(n)}(s-v)} b(s, T) ds dv \right) \\ &= \alpha \lambda_\infty \sum_{k=1}^n m_k^{(n)} \left(\int_t^T b(v, T) dv - b_k^{(n)} \int_t^T \frac{1}{b_k^{(n)}} \left(1 - e^{-b_k^{(n)}(s-t)} \right) b(s, T) ds \right) \\ &= \alpha \lambda_\infty \sum_{k=1}^n m_k^{(n)} \left(\int_t^T e^{-b_k^{(n)}(v-t)} b(v, T) dv \right). \end{aligned}$$

□

4.B.5 Proof of Corollary 4.1.

Part 1. If a partition $\mathcal{E}^{(n)}$ satisfies assumptions (i)-(iii), we first show for $t \geq 0$ that $\left(\lambda_t^{(n)}, N_t^{(n)}, L_t^{(n)}, \left(Y_t^{(k)} \right)_{k=1, \dots, n} \right)$ converges to $(\lambda_t, N_t, L_t, (Y_t^\omega)_{\omega \in \mathbb{R}^+})$ almost surely when $n \rightarrow \infty$. The almost sure convergence $\lim_{n \rightarrow \infty} \lambda_t^{(n)} = \lambda_t$ has been shown in p.150 (4.3.12) since we have $\lambda_t^{(n)} = \lambda_0 + \int_0^t Y_s^\omega \mu^{(n)}(d\omega)$ and since the function $\omega \in \mathbb{R}^+ \mapsto Y_t^\omega(w')$ is integrable with respect to the measure μ for fixed $w' \in \Omega$ as a sum of negative exponential functions. It follows that $N_t^{(n)} = \int_0^t \lambda_u^{(n)} du$ converges almost surely to N_t and $L_t^{(n)}$ to L_t , when $n \rightarrow \infty$. Finally, from (4.3.7), we have that $Y_t^{(k)} := Y_t^{b_k^{(n)}} = \int_0^t e^{-b_k^{(n)}(t-u)} \left(\alpha(\lambda_\infty - \lambda_u^{(n)}) du + \eta dL_u^{(n)} \right)$ for $k \in \{1, \dots, n\}$. If the partition $\mathcal{E}^n$ satisfies assumption (i)-(iii), the barycenter $b_k^{(n)}$ tends to be an element of $\mathbb{R}_+$, when $n \rightarrow \infty$. Hence, the sequence $(Y_t^{(k)})_{k=1, \dots, n}$ converges almost surely to $(Y_t^\omega)_{\omega \in \mathbb{R}^+}$.

Part 2. The dominated convergence theorem then provides

$$\mathbb{E}_t \left[e^{u_\lambda \lambda_T + u_L L_T + u_N N_T} \right] = \lim_{n \rightarrow \infty} \mathbb{E}_t \left[e^{u_\lambda \lambda_T^{(n)} + u_L L_T^{(n)} + u_N N_T^{(n)}} \right].$$

Moreover, for deriving the coefficient function $b(t, T)$, we first have from equation (4.3.2) that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n m_k^{(n)} = \int_0^\infty \mu(d\omega) = f_\gamma(0+) = 1. \quad (4.B.8)$$

We note that this quantity diverges toward infinity for the power-law and dML kernels, preventing from using this discretization scheme 4.3.3 in such cases. As in Part 1 of the proof, we then have almost surely in the limit $n \rightarrow \infty$:

$$\begin{aligned} \frac{\partial b(t, T)}{\partial t} &= \alpha \left(b(t, T) - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} b(v, T) dv \right) \omega \mu(d\omega) \right) \\ &\quad - \theta \left(b(t, T) \eta + u' \cdot (1, 0)^\top - \eta \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} b(v, T) dv \right) \omega \mu(d\omega) \right) e^{u' \cdot (0, 1)^\top} + 1. \end{aligned}$$

Then, from equation (4.3.2), we have

$$\begin{aligned} \frac{\partial b(t, T)}{\partial t} &= \alpha \left(b(t, T) - \int_t^T \frac{\partial f_\gamma(v-t)}{\partial t} b(v, T) dv \right) \\ &\quad - \theta \left(b(t, T) \eta + u' \cdot (1, 0)^\top - \eta \int_t^T \frac{\partial f_\gamma(v-t)}{\partial t} b(v, T) dv \right) e^{u' \cdot (0, 1)^\top} + 1, \end{aligned}$$

and finally from Leibniz integral rule to

$$\begin{aligned} \frac{\partial b(t, T)}{\partial t} &= -\alpha \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(v, T) dv \\ &\quad - \theta \left(u' \cdot (1, 0)^\top - \eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(v, T) dv \right) e^{u' \cdot (0, 1)^\top} + 1. \end{aligned}$$

□

4.B.6 Proof of Corollary 4.2.

By differentiating the conditional transform (4.3.14) with respect to u_λ, u_L, u_N and evaluating the sum of derivatives at $u = (0, 0, 0)$, we find the following conditional expectation

$$\begin{aligned} \mathbb{E}_t[w \cdot X_T] &= \alpha \lambda_\infty \int_t^T f_\gamma(v-t) b'(0, v, T) dv + b'(0, t, T) \lambda_t + w' \cdot J_t \\ &\quad - \int_0^\infty \left(\int_t^T e^{-\omega(v-t)} b'(0, v, T) dv \right) \omega Y_t^\omega \mu(d\omega), \end{aligned}$$

where $b'(0, t, T)$ is the sum of derivatives of the coefficient function $b(u', t, T)$ in

(4.3.15) with respect to (u_L, u_N) evaluated at $(0, 0)$, and which satisfies

$$\begin{aligned} \frac{\partial b'(0, t, T)}{\partial t} = & -\alpha \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv - \theta \left(-\eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(0, v, T) dv \right) w' \cdot (0, 1)^\top \\ & - \theta' \left(-\eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b(0, v, T) dv \right) \left(w' \cdot (1, 0)^\top - \eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv \right), \end{aligned}$$

with boundary condition $b'(0, T, T) = w_\lambda$ and where $\theta'(\xi) = \int o e^{\xi o} d\nu(o)$ is the derivative of the jump transform. Since $u = (0, 0, 0)$, we can choose $b(0, t, T) = 0$, $a(0, t, T) = 0$ and $c_k(0, t, T) = 0$ thanks to equation (4.3.13) and hence, the previous equation simplifies to

$$\begin{aligned} \frac{\partial b'(0, t, T)}{\partial t} = & -\alpha \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv - \theta(0) w' \cdot (0, 1)^\top \\ & - \theta'(0) \left(w' \cdot (1, 0)^\top - \eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv \right). \end{aligned}$$

We then have

$$\begin{aligned} \frac{\partial b'(0, t, T)}{\partial t} = & -\alpha \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv + \chi \eta \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv - w' \cdot (\chi, 1)^\top \\ = & (\chi \eta - \alpha) \frac{\partial}{\partial t} \int_t^T f_\gamma(v-t) b'(0, v, T) dv - w' \cdot (\chi, 1)^\top, \end{aligned} \quad (4.B.9)$$

where $\chi = \int o d\nu(o)$, which finally leads to

$$b'(0, t, T) = w_\lambda + (\chi \eta - \alpha) \int_t^T f_\gamma(v-t) b'(0, v, T) dv + (T-t) w' \cdot (\chi, 1)^\top.$$

□

4.B.7 Poof of Proposition 4.6

Proof. As for Proposition 4.5, we start by studying the discretized version of the conditional intensity (4.3.18) based on a partition $\mathcal{E}^n$ of size n , and hence obtain the following dynamics

$$dY_t^{(k,i)} = \left(-b_k^{(n)} Y_t^{(k,i)} + \alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda_t^{(n,i)} \right) \right) dt + \sum_{j=1}^d \eta^{(i,j)} O^{(j)} dN_t^{(n,j)} + \sigma^{(i)} \sqrt{\lambda_t^{(n,i)}} dW_t^{(i)}, \quad (4.B.10)$$

where $Y^{(k,i)} := Y^{b_k^{(n)},(i)}$ for $i = 1, \dots, d$, $k = 1, \dots, n$ using (4.3.19), and

$$d\lambda_t^{(n,i)} = \sum_{k=1}^n m_k^{(n,i)} dY_t^{(k,i)}, \quad (4.B.11)$$

where $m_k^{(n,i)} = \int_{\xi_{k-1}^{(n)}}^{\xi_k^{(n)}} \mu^{(i)}(d\omega)$, where $N^{(n,i)}$ is a counting process of intensity $\lambda^{(n,i)}$, and where $L_t^{(n,i)} = \sum_{j=1}^{N_t^{(n,i)}} O_j$. The infinitesimal generator of $Z^{(n)} = (\lambda^{(n)}, J^{(n)})^\top$ is given by

$$\begin{aligned} \mathcal{L}f(z) = & \sum_{i=1}^d \left\{ \frac{\partial f}{\partial \lambda^{(n,i)}}(z) \sum_{k=1}^n m_k^{(n,i)} \left(\alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda^{(n,i)} \right) - b_k^{(n)} Y^{(k,i)} \right) + \sum_{k=1}^n \frac{\partial f}{\partial Y^{(k,i)}}(z) \left[\alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda^{(n,i)} \right) \right. \right. \\ & \left. \left. - b_k^{(n)} Y^{(k,i)} \right] + \frac{1}{2} \lambda^{(n,i)} \sigma^{(i)2} \left(\sum_{k=1}^n m_k^{(n,i)} \right)^2 \frac{\partial^2 f}{\partial \lambda^{(n,i)2}}(z) + \frac{1}{2} \lambda^{(n,i)} \sigma^{(i)2} \sum_{k_1=1, k_2=1}^n \frac{\partial^2 f}{\partial Y^{(k_1,i)} \partial Y^{(k_2,i)}}(z) \right. \\ & \left. + \lambda^{(n,i)} \int_0^\infty \left[f \left(\left(\lambda^{(n,j)} + \eta^{(j,i)} o \sum_{k=1}^n m_k^{(n,j)} \right)_{j=1, \dots, d}, \left(Y^{(k,j)} + \eta^{(j,i)} o \right)_{k=1, \dots, n}, L^{(n,i)} + o, N^{(n,i)} + 1 \right) \right. \right. \\ & \left. \left. - f(z) \right] \nu^{(i)}(do) + \lambda^{(n,i)} \sigma^{(i)2} \left(\sum_{k=1}^n m_k^{(n,i)} \right) \sum_{k=1}^n \frac{\partial^2 f}{\partial \lambda^{(n,i)} \partial Y^{(k,i)}}(z) \right\}, \end{aligned}$$

with $z = Z_t^{(n)} = (\lambda_t^{(n)}, J_t^{(n)})^\top$ and for f a function in the domain of the infinitesimal generator of $Z^{(n)}$ with continuous partial derivatives. We then express the conditional transform as

$$f \left(t, T, Z_t^{(n)}, \left(Y_t^{(k,i)} \right)_{\substack{k=1, \dots, n \\ i=1, \dots, d}} \right) := \mathbb{E}_t \left[e^{u \cdot Z_T^{(n)}} \right] = \mathbb{E}_t \left[e^{u_\lambda \cdot \top \lambda_T^{(n)} + u_L \cdot \top L_T^{(n)} + u_N \cdot \top N_T^{(n)}} \right].$$

Since f must satisfy the PIDE (4.2.11) with boundary condition $f(T, T, Z_T^{(n)}) = e^{u_\lambda \cdot \top \lambda_T^{(n)} + u_L \cdot \top L_T^{(n)} + u_N \cdot \top N_T^{(n)}}$, we have that the transform of $Z^{(n)}$ is given by :

$$\begin{aligned} & f \left(t, T, Z_t^{(n)}, \left(Y_t^{(k,i)} \right)_{\substack{k=1, \dots, n \\ i=1, \dots, d}} \right) \\ &= \exp \left(\alpha(t, T) + \beta(t, T) \cdot \top \lambda_t^{(n)} + u_L \cdot \top L_t^{(n)} + u_N \cdot \top N_t^{(n)} + \sum_{i=1}^d \sum_{k=1}^n c_k^{(i)}(t, T) Y_t^{(k,i)} \right), \end{aligned}$$

with boundary conditions $\alpha(T, T) = 0$, $\beta(T, T) = u_\lambda$ and $c_k^{(i)}(T, T) = 0$, $\forall i \in 1, \dots, d$, $\forall k \in 1, \dots, n$. We hence find

$$\begin{aligned} 0 = & \frac{\partial \alpha(t, T)}{\partial t} + \sum_{i=1}^d \left\{ \frac{\partial \beta^{(i)}(t, T)}{\partial t} \lambda^{(n,i)} + \sum_{k=1}^n \frac{\partial c_k^{(i)}(t, T)}{\partial t} Y^{(k,i)} \right. \\ & + \beta^{(i)}(t, T) \sum_{k=1}^n m_k^{(n,i)} \left(\alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda^{(n,i)} \right) - b_k^{(n)} Y^{(k,i)} \right) + \sum_{k=1}^n c_k^{(i)}(t, T) \left(\alpha^{(i)} \left(\lambda_\infty^{(i)} - \lambda^{(n,i)} \right) - b_k^{(n)} Y^{(k,i)} \right) \\ & + \lambda^{(n,i)} \left[\left(\int \exp \left(\sum_{j=1}^d \beta^{(j)}(t, T) \eta^{(j,i)} o \sum_{k=1}^n m_k^{(n,j)} + u_{L,i} o + \sum_{j=1}^d \sum_{k=1}^n c_k^{(j)}(t, T) \eta^{(j,i)} o \right) \nu^{(i)}(do) \right) e^{u_{N,i}} - 1 \right] \\ & \left. + \frac{1}{2} \lambda^{(n,i)} \sigma^{(i)2} \left(\sum_{k=1}^n \left(m_k^{(n,i)} \beta^{(i)}(t, T) + c_k^{(i)}(t, T) \right) \right)^2 \right\}. \end{aligned}$$

From the previous equation, we can obtain the following system of equations

$$\begin{aligned}\frac{\partial \alpha(t, T)}{\partial t} &= - \sum_{i=1}^d \left(\alpha^{(i)} \lambda_{\infty}^{(i)} \sum_{k=1}^n \left(\beta^{(i)}(t, T) m_k^{(n, i)} + c_k^{(i)}(t, T) \right) \right), \\ \frac{\partial c_k^{(i)}(t, T)}{\partial t} &= \beta^{(i)}(t, T) m_k^{(n, i)} b_k^{(n)} + c_k^{(i)}(t, T) b_k^{(n)}, \quad \forall i = 1, \dots, d, \quad \forall k = 1, \dots, n, \\ \frac{\partial \beta^{(i)}(t, T)}{\partial t} &= \alpha^{(i)} \sum_{k=1}^n \left(m_k^{(n, i)} \beta^{(i)}(t, T) + c_k^{(i)}(t, T) \right) - \frac{1}{2} \sigma^{(i)2} \left(\sum_{k=1}^n \left(m_k^{(n, i)} \beta^{(i)}(t, T) + c_k^{(i)}(t, T) \right) \right)^2 \\ &\quad - \theta^{(i)} \left(\sum_{j=1}^d \eta^{(j, i)} \sum_{k=1}^n \left(\beta^{(j)}(t, T) m_k^{(n, j)} + c_k^{(j)}(t, T) \right) + u_{L, i} \right) e^{u_{N, i}} + 1, \quad \forall i = 1, \dots, d.\end{aligned}$$

This finally leads for the coefficients $\beta^{(i)}(\cdot)$ and $c_k^{(i)}(\cdot)$ to

$$c_k^{(i)}(t, T) = -b_k^{(n)} m_k^{(n, i)} \int_t^T e^{-b_k^{(n)}(v-t)} \beta^{(i)}(v, T) dv, \quad \forall i = 1, \dots, d, \quad \forall k = 1, \dots, n,$$

and

$$\begin{aligned}\alpha(t, T) &= \sum_{i=1}^d \left(\alpha^{(i)} \lambda_{\infty}^{(i)} \sum_{k=1}^n m_k^{(n, i)} \left(\int_t^T \beta^{(i)}(v, T) dv - b_k^{(n)} \int_t^T \int_v^T e^{-b_k^{(n)}(s-v)} \beta^{(i)}(s, T) ds dv \right) \right) \\ &= \sum_{i=1}^d \left(\alpha^{(i)} \lambda_{\infty}^{(i)} \sum_{k=1}^n m_k^{(n, i)} \left(\int_t^T e^{-b_k^{(n)}(v-t)} \beta^{(i)}(v, T) dv \right) \right).\end{aligned}$$

From Proposition 4.4 and Corollary 4.1 along with dominated convergence, passing to the limit $n \rightarrow \infty$ leads to the announced result. In particular, we have almost surely that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \left(m_k^{(n, i)} \beta^{(i)}(t, T) + c_k^{(i)}(t, T) \right) = - \frac{\partial}{\partial t} \int_t^T f_{\gamma^{(i)}}(v-t) \beta^{(i)}(v, T) dv, \quad \forall i \in 1, \dots, d.$$

□

Part III

Optimal control in finance

CHAPTER 5

Optimal liquidation under indirect price impact with propagator

At last, the third part of the thesis delves into stochastic optimal control in finance. We specifically focus on optimal liquidation problems as they are built upon price impact models and thus consider liquidity modeling from an intermediate/intraday time scale. We then propose in Chapter 5 a new framework of optimal execution strategies for a trader seeking to liquidate his large inventory based on a jump-dependent price impact model with propagator. This new jump-dependent price impact model best reproduces the empirical direct and indirect effects of market orders on the transaction price. More precisely, different choices of propagators are proposed and their implications in terms of temporary, permanent and transient impacts on the transaction price are discussed. For each choice of such kernels, we formulate the most relevant optimal liquidation problem faced by the trader, derive explicitly the related Hamilton-Jacobi-Bellman equation and solve it numerically. We then also show how our price impact model can be extended to incorporate the use of limit orders by the liquidating trader. Therefore, in Chapter 5, we aim to propose an alternative, realistic and flexible description of the order book's dynamics and liquidity, thereby contributing to bridging the gap between high-frequency liquidity modeling and optimal execution problems at intraday time scales.

Dupret, J. L., & Hainaut, D. Optimal liquidation under indirect price impact with propagator. *Available at SSRN:*
<http://dx.doi.org/10.2139/ssrn.4625578>

5.1 Introduction

A classical problem in finance is how an agent can sell (or buy) a large amount of shares and yet minimize adverse price movements generated by her own trades. *Large* means here that the amount the agent is interested in selling is too big to execute in a single trade. The agent must then take into account the balance between price impact (trade quicker) and price risk (take longer to complete all trades). Indeed, if an agent trades quickly, then her orders will walk through the buy side of the limit order book (LOB) and she will obtain lower prices for her orders, inducing a *temporary price impact*. Even if she splits each order into smaller trades so that each one does not walk the book and executes them rapidly in the market, then other traders will notice an excess of sell orders (or net order flow imbalance) and will adjust their quotes inducing again a negative price impact, which is in this case *transient* and/or *permanent*. On the other hand, if she trades slowly so as to avoid this price impact, then she will face the uncertainty of future prices, which could have dropped to a less desirable level by the time the agent executes the whole order. The aim of optimal liquidation is hence to propose an optimal strategy such as to balance these two factors (liquidity vs. uncertainty), based on a coherent price impact model reproducing the underlying asset price and the impact of the agent's transactions on this price. We refer to Gökay et al. (2011), Cartea et al. (2015), Guéant (2016), Bouchaud et al. (2018), Donnelly (2022) for a review of the most important results concerning optimal execution problems.

Typically, when modeling the price of an asset, three different time scales can be considered, see the general Introduction. At low frequencies, the dynamic of stock prices can often be reproduced by a diffusive process. At the other end, when dealing with very high frequencies, empirical features and statistical properties of the LOB dynamics have to be modeled (long-range dependence of trades, volatility clustering, high-frequency resilience of the price, market liquidity, etc.). At such very high frequencies, one then has to describe precisely the LOB dynamic. A variety of such models have been proposed in Abergel and Jedidi (2013), Cont and De Larrard (2013), Gareche et al. (2013), Huang et al. (2015) and especially in Bacry and Muzy (2014) with a tick-by-tick price model ruled by Hawkes processes, which describes jointly the order flow and the price dynamics and which reproduces numerous empirical properties of high-frequency markets. In between of low and high-frequency models, price impact models consider an intra-day intermediate time scale, anywhere from seconds to hours. They tend to discard most of the LOB events such as limit orders, cancellations, market orders, etc. and instead concentrate on describing the price impact of the agents' transactions. Their goal is to be more tractable than high-frequency models and to offer analytical results on practical issues such as the optimal liquidation problems studied in this paper. The usual setup is well described in Gatheral (2010) where the execution price S^ν

of a transaction is given by

$$S_t^\nu = S_0 + \sigma W_t + \int_0^t f(\nu_s)G(t-s)ds, \quad t \in [0, T], \quad (5.1.1)$$

with $T > 0$ the trading horizon, $S_0 > 0$ the initial price, ν_s the rate of trading of the liquidating agent at time s , $f(\nu)$ the instantaneous price impact of an agent trading at speed ν , G the *decay kernel* or *propagator* and W a noise process (usually a Brownian motion). As explained in Alfonsi and Blanc (2016), $f(\nu)G(0)$ captures the temporary/instantaneous impact of trades on the price, $f(\nu)G(+\infty)$ the permanent impact and $f(\nu)[G(0+) - G(+\infty)]$ the transient impact. That is why the two extreme cases where G is a Dirac's delta $\delta(t-s)$ on $[0, T]$ and when $G \equiv 1$ are referred in the literature as temporary price impact and permanent price impact, respectively. They are core features in the well-known model of Almgren and Chriss (1999) and Almgren and Chriss (2001), up to a multiplicative constant. Moreover, the price impact model (5.1.1) also allows to generalize most of the standard processes considered in the optimal liquidation literature, see Bertsimas and Lo (1998), Bouchaud et al. (2003), Almgren et al. (2005), Alfonsi et al. (2010), Forsyth et al. (2012), Obizhaeva and Wang (2013) and Donnelly (2022). Since we are interested in this paper in optimal execution problems, we will then work with such price impact models observed at intermediate frequency. However, we aim at expanding these models by including more specific events and features of the LOB, as we will show with Figures 5.1.1 and 5.1.2. One of our main goal is then to contribute to bridging this gap between high-frequency price models and the optimal liquidation framework by proposing a tractable model reproducing some empirical evidence of market liquidity at high frequency. We refer to Alfonsi and Blanc (2016) or Cartea et al. (2014) who were among the first to propose such mixed-framework models incorporating both high-frequency and middle-frequency feature of the LOB.

As mentioned above and explained in Gatheral (2010), Alfonsi and Blanc (2016) and Donnelly (2022), we have that price impact can be temporary, transient and/or permanent. Market orders (MOs) that walk the LOB have a price impact because they obtain average execution prices worse than the best quote and this price impact is temporary because it is assumed that the LOB will replenish very quickly and, after restocking, there is no effect on the midprice of the asset. On the other hand, investors' trading activity has a transient and permanent effect on midprices since the traders' order flow (*i.e.* the number and the size of MOs) conveys information that is impounded on the midprice of the asset, as investigated in the study of Cont et al. (2014) on order flow imbalance. In general, when there is positive net order flow (more buy than sell volume) the midprice tends to drift up, and when there is negative net order flow (more sell than buy volume) the midprice tends to drift down. Thus, when an investor is executing a large market order, this adds one-sided pressure to the market's order flow and hence on midprices: up if the investor is buying or down if selling. In this paper, this downward permanent

(and transient) pressure on the price will be modeled by an increase in the mean-reverting intensity of a point process with negative jump size, which will allow to convey information about MOs' permanent and transient price impact. This jump process will replace the term $f(\nu_s)ds$ in the standard setting (5.1.1) as we will see in Section 5.2 with equation (5.2.3). We will also investigate various choices of decay kernel G so as to combine different ways of modeling these temporary, transient and permanent price impacts.

More precisely, at any one time, we observe empirically that the number of shares displayed and available in the market at the quoted bid/ask is limited, see Figure 5.1.1 and 5.1.2 below. The liquidating trader will then split its sell order in different trade sizes, some of which could be very small at a particular time compared to the depth of the LOB. If a market order at time t does not walk the LOB, *i.e.* the size of this transaction is smaller than the depth at the best bid, there will not be any direct change in the stock's midprice (*no temporary direct price impact from walking the LOB*), even if this will lead to more pressure towards a price decrease (*indirect permanent and transient impact from the net order flow imbalance*). As stated above, this downward pressure in the price will be modeled in the sequel by an increase in the mean-reverting intensity of a jump process with negative jump size. Including a jump process for representing MO's price impact makes sense since the depth at the best bid is random/unpredictable at any time t : most of the time a small sell MO will not lead to a direct shift in price (no jumps) but for whatever reason, if the quantity available at the best bid is really low at this time, this small order of the trader will however walk the LOB and lead to an immediate shift (jump) in the midprice. Similarly, if the LOB is particularly deep at that time t , even a large market sell order could not lead to a direct shift in the midprice (*no temporary direct price impact*) but it will however put a lot of downwards pressure in the price (*indirect permanent/transient price impact*). This will hence translate in our model by a strong increase in the intensity of the jump process. These features of the LOB are confirmed in Cartea et al. (2015), Cartea and Jaimungal (2016), Bouchaud et al. (2018) and especially in Bacry and Muzy (2014) where a 4-dimensional Hawkes process is used to reproduce these direct and indirect price impacts. These features are also confirmed empirically by the following Figure 5.1.1, which represents the LOB for the AMZN share on June 21, 2012 at 10:36:47.016 and the related temporary price impact of a sell MO in function of volume. Similar figures and results are depicted in Cartea and Jaimungal (2016) for the SMH stock price.

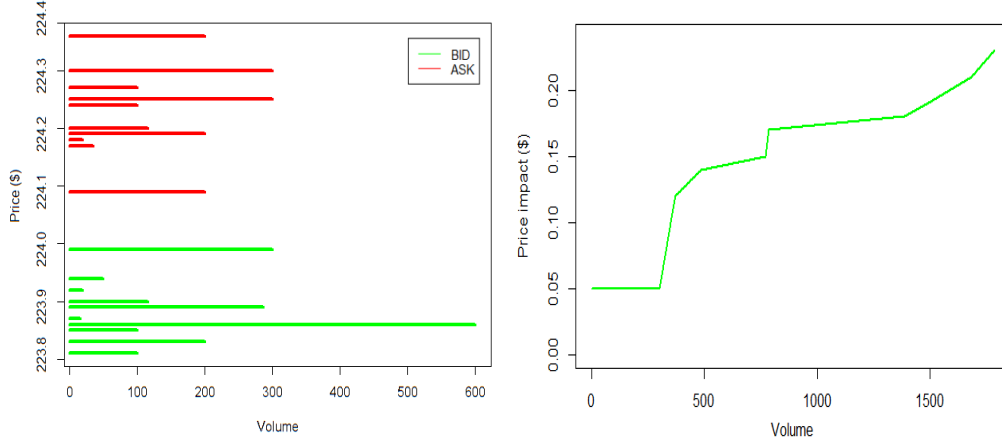


Figure 5.1.1: Limit order book of AMZN on the 21/06/2012 at 10:36:47.016 and related (interpolated) temporary price impact of a market sell order in function of the volume sold. In green, bid prices and in red, ask prices.

We clearly see a plateau in the temporary price impact curve for small volumes, meaning that such small MOs do not walk the LOB (sell volume $<$ depth at best bid) and hence do not lead to any direct temporary price impact. Indeed for sell MOs with a volume below 300, there is no direct temporary price impact besides the bid-ask spread. This is confirmed by the following Figure 5.1.2 where multiple temporary price impact curves in function of the volume sold are depicted inside the same minute for the AMZN stock price.

In each price impact curve of Figure 5.1.2, we again see plateaus for small volumes, translating the fact that small enough sell MOs (with volume lower than the depth at the best bid) do not lead to any direct temporary price impact besides the bid-ask spread. More importantly, we see in both plots that even the average temporary impact curve over the minute (bold line) remains constant for sell volumes below 100. Moreover, once the volume sold is large enough for triggering a price impact, this temporary impact seems to be best reproduced by a power-law function with exponent below 1 and more precisely between 0.4 and 0.8. This is consistent with the literature and the so-called square-root law as investigated in Torre (1997), Bouchaud et al. (2003), Almgren et al. (2005) and Gatheral (2010). Such behaviors of the temporary impact curves justify the use of an indirect price impact model based on a jump process, as we will introduce in the following Section 5.2. This point process will indeed jump whenever the sell order is sufficiently large (here above 100) and the size of the negative jump will then scale as a power-law function of the volume sold. As mentioned above, the intensity of this jump process

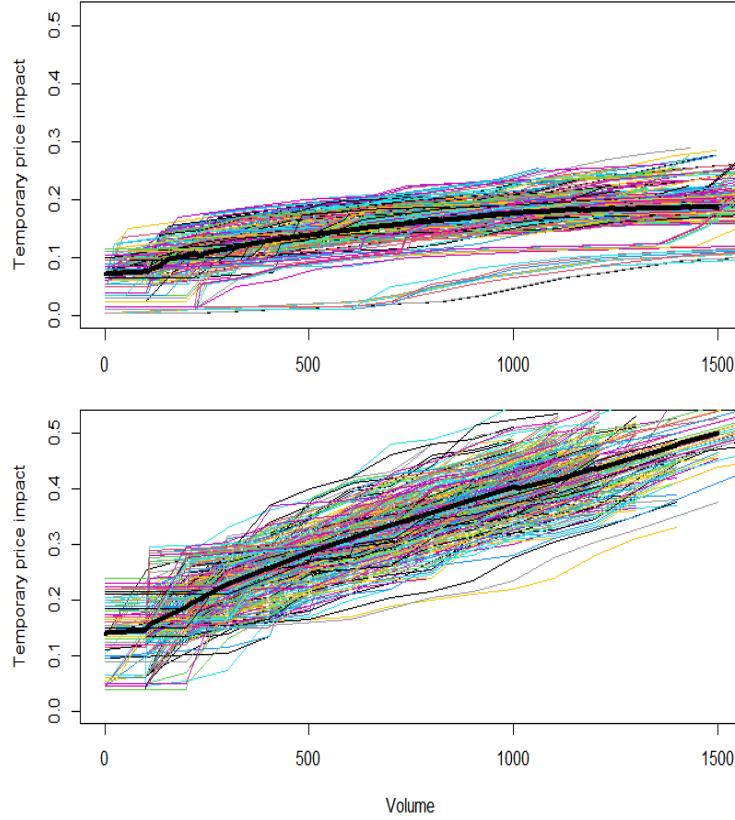


Figure 5.1.2: Temporary price impact in function of volume for AMZN on Jun 21, 2012 for every millisecond between 10:36 and 10:37 am (upper plot) and between 09:43 and 09:44 am (lower plot). In black wider line is depicted the mean temporary price impact over the corresponding minute.

will also depend on the volume of the sell MO as each sell order puts a permanent downwards pressure on the price through the net order flow imbalance, even if there is no direct temporary impact (no jump). Very few price impact models incorporate this absence of temporary negative impact for sufficiently low volumes, see *e.g.* the discussion in Section 2.1.3 of Bacry and Muzy (2014). However, being able to trade with MOs without having any temporary price impact for small enough sell orders has a strong impact on the optimal strategy used by the liquidating trader. Moreover, we will see that this model extends the stochastic liquidity factor of Fruth et al. (2014) and Fruth et al. (2019) in the form of a trade-driven intensity process.

Finally, we raise the point that optimal liquidation problems based on jump processes are already studied extensively in the literature, see *e.g.* Bacry and Muzy (2014), Bayraktar and Ludkovski (2014), Cartea et al. (2014), Guéant and Lehalle (2015), Da Fonseca and Malevergne (2021), Sadoghi and Vecer (2022), or Chapters 8.1-8.3 of Cartea et al. (2015). However, the jump processes in these papers are systematically used to model the occurrence of market and limit orders sent by the liquidating trader. In our setting, the jump process instead allows to model explicitly the occurrence of the temporary/permanent/transient price impacts arising from continuous trading with MOs since all trades do not lead to a direct negative impact on the execution price, see Figure 5.1.2. Moreover, we propose in Section 5.3 an extension of our price impact model so as to include strategies using both market and limit orders (based on a second jump process). This will hence provide a more complete and realistic price impact model, allowing for a better modeling of the LOB dynamic.

The contributions of this paper are therefore multiple. We first introduce a new jump-dependent price impact model with propagator allowing to best reproduce the empirical direct and indirect effects of market orders on the transaction price, as described above and in Figure 5.1.2. We then discuss in Section 5.2 various ways of modeling in our setting the temporary, permanent and/or transient price impact of trades based on different choices of propagators G . For each choice of such kernels, we formulate the most relevant optimal liquidation problem faced by the agent, derive explicitly the related HJB equation and solve it numerically. This allows to obtain and compare the optimal trading strategies in each scenario and to discuss the implications of each model (and its parameters) on the optimal trading rate of the liquidating agent. We finally extend in Section 5.3 the initial price impact model based solely on market orders such as to incorporate the possibility for the agent to use both market and limit orders. We hence provide with this paper a new jump-dependent price impact model with stochastic liquidity factor, leading to a more realistic description of the underlying LOB dynamic and including specific high-frequency features of the order book.

The remainder of this chapter is organized as follows. Section 5.2 introduces the indirect price impact model with propagator. Sections 5.2.1, 5.2.2 and 5.2.3 consider various assumptions on the propagator G with, respectively, a temporary, a permanent and a transient form for G . Section 5.3 generalizes the model by including limit orders. Section 5.4 concludes.

5.2 The indirect price impact model with propagator

We introduce the following liquidation model in continuous-time for a selling agent (which is also applicable to buying agents by symmetry). We assume at first that only MOs are posted by the trader (limit orders are then discussed in Section 5.3).

Moreover, no bid-ask spread is taken into account in what follows but the model can be easily extended so as to include it. We first define the following quantities:

- T is the time horizon required for liquidating the shares, *i.e.* the trading horizon.
- $(\nu_t)_t := (\nu_t)_{0 \leq t \leq T}$ is the trading rate process, *i.e.* the (non-negative) speed at which the agent is liquidating shares using MOs and which can be seen as the volume sold on each infinitesimal step of time. We denote by $(\nu_t)_t$ the whole strategy/process on $[0, T]$ and by ν a particular value in $\mathbb{R}^+$ of this strategy.
- $Q^\nu := (Q_t^\nu)_{0 \leq t \leq T}$ is the agent's inventory of remaining shares to liquidate, affected by how fast she trades and given by $dQ_t^\nu = -\nu_t dt$ with $Q_0^\nu = Q > 0$. We do not assume here that the agent liquidates all the shares during the horizon T . Hence, it could be that $Q_T^\nu \geq 0$, *i.e.* the agent falls short of her target to liquidate all the shares before T but she incurs in this case a terminal penalty (whose form is detailed below). However, we do assume that the agent cannot liquidate more than the Q initial shares such that we cannot have $Q_T^\nu < 0$ (as implied below by the stopping time τ in (5.2.1)).
- $S^\nu := (S_t^\nu)_{0 \leq t \leq T}$ is the execution price affected by the chosen trading rate.
- $X^\nu := (X_t^\nu)_{0 \leq t \leq T}$ is the agent's cash/wealth process from her execution strategy, defined by $X_t^\nu := \int_0^t S_s^\nu \nu_s ds$.

We fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ where T is the deterministic trading horizon. We then need to define the stopping time τ as

$$\tau := T \wedge \inf\{t \geq 0 : Q_t^\nu = 0\}, \quad (5.2.1)$$

which is the minimum of the maturity T and the first time the inventory hits zero (because then no more trading is necessary), see *e.g.* Chapter 7.2, 7.4, 8.4 of Cartea et al. (2015). This stopping time τ is therefore defined as the last trading time of the agent (after which the wealth X does not vary anymore). The admissible set $\mathcal{A}$ of trading strategies consists in non-negative, $\mathbb{F}$ -predictable and uniformly bounded from above on $[0, \tau)$ trading rates $(\nu_t)_t$:

$$\mathcal{A} = \{(\nu_t)_t \mid (\nu_t)_t \geq 0, \mathbb{F}\text{-predictable}, (\nu_t)_t \in L^\infty(\Omega \times [0, \tau))\}. \quad (5.2.2)$$

Therefore, the strategy $(\nu_t)_t$ takes its values ν in some bounded set $\Gamma \subset \mathbb{R}^+$ on $[0, \tau)$. As an extension of the price impact model of Gatheral (2010), we further assume that the execution price S_t^ν at time $t \leq \tau$ follows

$$S_t^\nu := S_t + \int_0^t G(t-s) dL_s^\nu, \quad (5.2.3)$$

where $S_t = S_0 + \sigma W_t$ is a continuous martingale representing the unaffected stock price and the decay kernel $G : [0, T] \rightarrow [0, +\infty]$ is such that G is a measurable

function, square-integrable on $[0, T]$ and $G(0) := \lim_{t \rightarrow 0} G(t)$ exists. We will consider throughout the paper a $\mathbb{F}$ -Brownian motion as the continuous martingale $W = (W_t)_{t \geq 0}$. We then further define the jump process for $t \in [0, \tau]$ by

$$L_t^\nu := - \int_0^t \kappa \nu_s^\gamma dN_s^\nu,$$

where the jumps' size depends on the size of the trading rate through the exponent $\gamma \geq 0$ and where $N^\nu := (N_t^\nu)_{0 \leq t \leq \tau}$ is a counting process independent from W with mean-reverting intensity¹

$$d\lambda_t^\nu = (\eta \nu_t - \alpha \lambda_t^\nu) dt \quad \text{such that} \quad \lambda_t^\nu = \lambda_0 e^{-\alpha t} + \eta \int_0^t e^{-\alpha(t-s)} \nu_s ds, \quad (5.2.4)$$

where $\lambda_0, \eta, \alpha \geq 0$. Since $(\nu_t)_t \in \mathcal{A}$ is non-negative, the intensity $\lambda^\nu = (\lambda_t^\nu)_{0 \leq t \leq \tau}$ is also non-negative and bounded from below by $\lambda_{\min} := \lambda_0 e^{-\alpha T}$. In practice, this means that purchasing shares is not allowed in our model at any time during the trading horizon. Moreover, this model induces that the execution price of a transaction depends both *indirectly* on the strategy $(\nu_t)_t$ chosen by the trader via the intensity λ^ν of N^ν , as well as *directly* on $(\nu_t)_t$ through the jump size $(-\kappa \nu_t^\gamma)$ of the point process L^ν and through the propagator G in (5.2.3). Indeed, high trading rates increase the probability of jumps and hence of a negative impact on the transaction price, which is then proportional to the size of that trade ν and scales more precisely as a power law function of ν with exponent $\gamma \leq 1$, as seen in Section 5.1 and Figure 5.1.2. Empirically, it makes sense that λ_t^ν depends on the past trading rates $(\nu_s)_{0 \leq s \leq t}$ since a sell MO puts a permanent and transient downward pressure on the price, which translates in our model by a higher probability of a negative jump, see again Section 5.1. Therefore, we always have in this model a *permanent and transient* indirect impact of MOs on the price through the mean-reverting intensity λ^ν , even if the decay kernel G ruling the direct price impact can either be temporary (*i.e.* formally a Dirac's delta, cf. Section 5.2.1) or take other forms (Sections 5.2.2 and 5.2.3). This allows to obtain a more flexible setting than many liquidation models proposed in the literature. Note that the mean-reversion property of the intensity (5.2.4) induces an (indirect) transient price impact in our model while the fact that the mean-reversion level of λ^ν depends on $(\nu_t)_t$ leads to a permanent impact of the trades. As argued in Gatheral (2010) or Gatheral and Schied (2013), considering the martingale $S_t = S_0 + \sigma W_t$ (with W Brownian motion) for the unaffected price then allows to implicitly incorporate the cumulative impact of others' trading (the trading crowd in the terminology of Huberman and Stanzl (2004)) and is quite standard in the literature. Indeed, negative prices only occur with negligible probability at the considered time scales. Moreover, the drift

¹One may consider instead as λ_0 a deterministic function of time so as to include time-of-day effects on the liquidity, as in Cont et al. (2014) and Fruth et al. (2014). Moreover, one might sometimes expect or observe in practice a decreasing intensity λ^ν w.r.t trading, leading in this case to $\eta \leq 0$.

term of S is usually omitted for such small time scales. Finally, the agent's cash process X^ν satisfies $dX_t^\nu = S_t^\nu \nu_t dt = \left(S_t + \int_0^t G(t-s) dL_s^\nu \right) \nu_t dt$ and therefore, the expected revenues from the sale up to time $t \leq \tau$ is denoted

$$\mathbb{E}[X_t^\nu] = \mathbb{E} \left[\int_0^t S_s^\nu \nu_s ds \right] = \mathbb{E} \left[\int_0^t \left(S_s + \int_0^s G(s-u) dL_u^\nu \right) \nu_s ds \right]. \quad (5.2.5)$$

It is important to note that the price impact model described in this section can be seen as a stochastic model for the liquidity factor, also known as Kyle's Lambda (1985), in the form of a trade-driven intensity process λ^ν playing the role of a stochastic prefactor to the impact model (up to a constant κ and the propagator G). We will indeed prove in different settings that the optimal policy does not depend on the price jumps but only on its intensity. By focusing on this stochastic liquidity effect, we can hence compare our results with the deterministic liquidity case of Fruth et al. (2014) and Curato et al. (2017) or the stochastic case of Fruth et al. (2019) and Graewe and Horst (2017). Now that the price impact model is introduced for a general kernel G , we specify various suitable choices of such propagators in (5.2.3) so as to best fit the empirical observations of Figures 5.1.1 and 5.1.2.

5.2.1 Temporary price impact kernel

We consider in this section only a temporary direct impact of trading on the execution price, meaning that a sell MO at a time t only has a direct effect on the price of this transaction but not on the price of subsequent transactions. Recall however that a permanent and transient impact of MOs is always included indirectly in our setting via the mean-reverting intensity λ^ν of the jump process in (5.2.4). In this temporary case, it is best to start by defining the dynamic of the wealth process X^ν at the stopping time τ (5.2.1) in the following way

$$X_\tau^\nu := \int_0^\tau S_s \nu_s ds - \int_0^\tau \kappa \nu_s^{\gamma+1} dN_s^\nu = \int_0^\tau S_s \nu_s ds - \sum_{j=1}^{N_\tau^\nu} \kappa \nu_{t_j}^{\gamma+1}. \quad (5.2.6)$$

We then find using integration by parts with $Q_\tau^\nu = Q_t^\nu - \int_t^\tau \nu_s ds$ and with $\tau \geq t$,

$$\int_t^\tau S_s \nu_s ds = S_t Q_t^\nu - S_\tau Q_\tau^\nu + \int_t^\tau \sigma Q_s^\nu dW_s.$$

Since Q^ν is bounded (and belongs more precisely to $[0, Q]$ in the interval $[0, \tau]$), the third term is a martingale and since $dX_t^\nu = S_t^\nu \nu_t dt$, we find

$$\mathbb{E}[X_\tau^\nu | \mathcal{F}_t] = X_t^\nu + S_t Q_t^\nu + \mathbb{E} \left[-S_\tau Q_\tau^\nu - \kappa \int_t^\tau \nu_s^{\gamma+1} \lambda_s^\nu ds \middle| \mathcal{F}_t \right]. \quad (5.2.7)$$

As $(\nu_t)_t \in \mathcal{A}$ is non-negative and $\mathbb{F}$ -predictable, see (5.2.2), we indeed have from Lu (2013) that $\mathbb{E} \left[\int_t^\tau \kappa \nu_s^{\gamma+1} dN_s^\nu \middle| \mathcal{F}_t \right] = \mathbb{E} \left[\int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds \middle| \mathcal{F}_t \right]$. This expression (5.2.6) for the agent's wealth appears to be the only reasonable way for modeling this temporary direct impact on the price in our setting. Indeed, since trading occurs continuously during the trading horizon and since the direct impact of a trade only affects the transaction price at this time and not the subsequent ones, we have that the wealth is simply the proceeds of the sale at the unaffected price S minus the sum of these negative impulsive impacts $\kappa \nu_{t_j}^\gamma$ (each time they occur) times the quantity sold ν_{t_j} at that time. As explained in Gatheral (2010) and in Abi Jaber and Neuman (2022), considering only a temporary direct price impact amounts to use formally a Dirac's delta distribution $G(t-s) = \delta(t-s)$ on $[0, T]$ as propagator in the execution price S^ν (5.2.3). We refer to Appendix 5.A where we show formally in the temporary case the equivalence between the wealth process (5.2.6) and the execution price (5.2.3) with a Dirac's delta kernel G .

Now suppose that the agent is willing to optimize the following performance criterion

$$H^\nu(t, x, S, q, \lambda) = \mathbb{E}_{t,x,S,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_u^\nu)^2 du \right], \quad (5.2.8)$$

where $\mathbb{E}_{t,x,S,q,\lambda}[\cdot]$ denotes the expectation conditional to $X_{t-}^\nu = x$, $S_t = S$, $Q_t^\nu = q$, $\lambda_t^\nu = \lambda$ and where τ is the stopping time defined in (5.2.1). The corresponding value function at $t \leq \tau$ is then

$$H(t, x, S, q, \lambda) = \sup_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t,x,S,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_u^\nu)^2 du \right], \quad (5.2.9)$$

where $\mathcal{A}$ is the admissible set (5.2.2) of trading strategies. We first see that if the agent's strategy reaches the terminal date T with inventory left, then she must execute a final order to sell the required number of shares Q for a total revenue at T of $Q_T^\nu (S_T - \pi \lambda_T^\nu Q_T^\nu)$, where π is the terminal liquidation penalty parameter² that includes the cost of walking the book at T and any other additional penalties that the agent must incur for the execution of the trade at the terminal date. The terminal penalty $Q_T^\nu (S_T - \pi \lambda_T^\nu Q_T^\nu)$ without the term λ_T^ν is classical in the literature when the price impact is given by the model (5.1.1) and does not depend on a jump process, see Chap. 6 of Cartea et al. (2015) and references therein. However, it is necessary in our context to adapt this terminal penalty. Indeed, depending on the shape of the LOB at maturity T , it could be that liquidating the Q_T^ν remaining shares using MOs does not lead to a direct negative price impact and hence, the occurrence of this negative terminal penalty must again depend on the jump process

²We consider here the unaffected price S_T and not S_T^ν since the direct impact is temporary in this section, the terminal penalty being imputed here in place of this temporary impact.

N^ν with intensity λ^ν . We again refer to Appendix 5.A for the formal derivation of this terminal penalty $Q_\tau^\nu(S_\tau - \pi\lambda_\tau^\nu Q_\tau^\nu)$ in the temporary case with Dirac propagator G . Secondly, the last term in (5.2.9) is a running inventory penalty, which is not a financial cost to the agent's strategy (and hence not depending on λ^ν) but which allows to incorporate the agent's urgency for executing the trade. It has been shown in Cartea et al. (2017) that this penalty is equivalent to a form of risk aversion from the agent. An alternative risk aversion penalty is also discussed in Appendix 5.B. Finally, as seen above in Figure 5.1.2 and as shown in Torre (1997), Almgren et al. (2005), Gatheral (2010) and Cartea et al. (2015), we have empirically that the *temporary* (or instantaneous) price impact depends on the square root of the trading rate ($\gamma = 1/2$) and more generally on the power-law ν^γ with $\gamma \leq 1$. Our setup in this section is in fact similar in terms of temporary impact as the one of Almgren (2003), Almgren et al. (2005), Gatheral (2010) where they also use $G(t-s) = \delta(t-s)$ but without jump process N^ν and based instead on equation (5.1.1) with $f(\nu) = -\kappa\nu^\gamma$ and $\gamma \in [0.4, 0.8]$.

5.2.1.1 Temporary impact with mean-reverting intensity

We first consider that the jump process N^ν is ruled by the mean-reverting intensity (5.2.4). We will then show in Section 5.2.1.2 that the optimization problem (5.2.9) can be simplified by choosing an alternative expression for λ^ν . Recall first that the set $\mathcal{A}$ of admissible trading rates $(\nu_t)_t$ consists in uniformly bounded on $[0, \tau)$, non-negative and $\mathbb{F}$ -predictable strategies. From equation (5.2.6) and from (5.2.7), we can then rewrite the expected wealth (5.2.5) in the following way at the last trading time τ :

$$\mathbb{E}_{t,x,S,q,\lambda}[X_\tau^\nu] = x + qS + \mathbb{E}_{t,S,q,\lambda} \left[-S_\tau Q_\tau^\nu - \int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds \right],$$

with $X_{t-} = x, S_t = S, Q_t^\nu = q, \lambda_t^\nu = \lambda$. Note that when considering formally a Dirac delta $G(t-s) = \delta(t-s)$ in equation (5.2.3) for the execution price S^ν and using a Fubini's theorem for distribution (Chap. 4, Sec. 3, Thm. IV. of Schwartz (1966)), we find equivalently from equation (5.2.5) that

$$\begin{aligned} \mathbb{E}_{t,x,S,q,\lambda}[X_\tau^\nu] &= x + \mathbb{E}_{t,S,q,\lambda} \left[\int_t^\tau S_s \nu_s ds - \int_t^\tau \left(\int_0^s G(s-u) \kappa \nu_u^\gamma \lambda_u^\nu du \right) \nu_s ds \right] \\ &= x + qS + \mathbb{E}_{t,S,q,\lambda} \left[-S_\tau Q_\tau^\nu - \int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds \right]. \end{aligned}$$

We can then rewrite the performance criterion (5.2.8) as

$$\begin{aligned} H^\nu(t, x, S, q, \lambda) &= \mathbb{E}_{t,x,S,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + qS + \mathbb{E}_{t,q,\lambda} \left[- \int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 - \phi \int_t^\tau (Q_s^\nu)^2 ds \right]. \end{aligned}$$

And we finally have for the value function (5.2.9) :

$$\begin{aligned}
H(t, x, S, q, \lambda) &= x + qS + \sup_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t, q, \lambda} \left[- \int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 - \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\
&= x + qS - \inf_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t, q, \lambda} \left[\int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds + \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 + \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\
&= x + qS - \inf_{(\nu_t)_t \in \mathcal{A}} h^\nu(t, q, \lambda) = x + qS - h(t, q, \lambda). \tag{5.2.10}
\end{aligned}$$

First, observe that $h(t, q, \lambda)$ is non-negative since $h^\nu(t, q, \lambda)$ is non-negative for every $(\nu_t)_t \in \mathcal{A}$. Secondly, since $(\nu_t)_t \in \mathcal{A}$ is bounded on $[0, \tau)$ (with values in Γ), we have from (5.2.4) that λ^ν is bounded on $[0, \tau)$ (and Q^ν bounded on $[0, \tau]$). Moreover, there always exists a $(\nu_t)_t \in \mathcal{A}$ such that $h^\nu(t, q, \lambda) < \infty$ and hence, $h(t, q, \lambda) := \inf_{(\nu_t)_t \in \mathcal{A}} h^\nu(t, q, \lambda) < \infty$ for all $(t, q, \lambda) \in \mathcal{S} := [0, T] \times [0, Q] \times [\lambda_{\min}, +\infty)$. Therefore, $H(t, x, S, q, \lambda)$ is also finite for any given values of $(t, q, \lambda, x, S) \in \mathcal{S} \times \mathbb{R} \times \mathbb{R}$. It can be derived from Graewe and Horst (2017) that $h(t, q, \lambda)$ is continuous on $\mathcal{S}$ and C^1 on $S := [0, T] \times (0, Q] \times [\lambda_{\min}, +\infty)$. Then, from Sennewald (2007), we can derive the following verification principle based on the Hamilton-Jacobi-Bellman (HJB) equation.

Proposition 5.1. *Let $\gamma > 0$ and a continuous function $J : \bar{\mathcal{S}} \rightarrow \mathbb{R}^+$ assumed to be $C^1(\mathcal{S})$. Let also an admissible Markovian³ $(\nu_t^*)_t$ strategy satisfying $\forall (t, q, \lambda) \in \mathcal{S}$:*

$$\partial_t J(t, q, \lambda) + \phi q^2 + (\eta \nu^* - \alpha \lambda) \partial_\lambda J(t, q, \lambda) - \nu^* \partial_q J(t, q, \lambda) + \kappa (\nu^*)^{\gamma+1} \lambda = 0, \tag{5.2.11}$$

as well as the first order condition

$$\partial_\nu \left[(\eta \nu - \alpha \lambda) \partial_\lambda J(t, q, \lambda) - \nu \partial_q J(t, q, \lambda) + \kappa \nu^{\gamma+1} \lambda \right] \Big|_{\nu=\nu^*} = 0, \tag{5.2.12}$$

and if furthermore $J(T, q, \lambda) = \pi \lambda q^2$ and $J(t, 0, \lambda) = 0$, then $(\nu_t^*)_t$ is an optimal strategy and J is the value function h .

Proof. We first show that for all $\nu = \nu(t, q, \lambda)$ and $\nu^* = \nu^*(t, q, \lambda) \in \Gamma$ with $(t, q, \lambda) \in \mathcal{S}$:

$$\begin{aligned}
&\partial_t J(t, q, \lambda) + \phi q^2 + (\eta \nu^* - \alpha \lambda) \partial_\lambda J(t, q, \lambda) - \nu^* \partial_q J(t, q, \lambda) + \kappa (\nu^*)^{\gamma+1} \lambda \\
&\leq \partial_t J(t, q, \lambda) + \phi q^2 + (\eta \nu - \alpha \lambda) \partial_\lambda J(t, q, \lambda) - \nu \partial_q J(t, q, \lambda) + \kappa \nu^{\gamma+1} \lambda,
\end{aligned}$$

and hence that

$$\eta \nu^* \partial_\lambda J(t, q, \lambda) - \nu^* \partial_q J(t, q, \lambda) + \kappa (\nu^*)^{\gamma+1} \lambda \leq \eta \nu \partial_\lambda J(t, q, \lambda) - \nu \partial_q J(t, q, \lambda) + \kappa \nu^{\gamma+1} \lambda. \tag{5.2.13}$$

³Since the strategy is here assumed to be Markov, we then denote $\nu = \nu(t, q, \lambda) := \nu_t(q, \lambda) \in \Gamma$ and $\nu^* = \nu^*(t, q, \lambda) := \nu_t^*(q, \lambda) \in \Gamma$ for any values of $(t, q, \lambda) \in \mathcal{S}$.

Indeed, from the first order condition (5.2.12), ν^* is a local minimum point of the RHS of (5.2.13). Moreover, the states Q^ν and λ^ν are linear in the control variable ν (see Section 5.2) and the running cost function in (5.2.10) given by $\kappa\nu^{\gamma+1}\lambda + \phi q^2$ is a non-negative differentiable and strictly convex function in ν since $\gamma > 0$. Therefore, the right-hand side $\eta\nu\partial_\lambda J(t, q, \lambda) - \nu\partial_q J(t, q, \lambda) + \kappa\nu^{\gamma+1}\lambda$ of (5.2.13) is strictly convex in ν and hence, $\nu^* = \nu^*(t, q, \lambda)$ is both a local and a global minimum point of (5.2.13), leading together with (5.2.11) to the classical HJB equation for all $(t, q, \lambda) \in \mathcal{S}$:

$$\partial_t J(t, q, \lambda) + \phi q^2 + \min_{\nu \in \Gamma} \{ (\eta\nu - \alpha\lambda)\partial_\lambda J(t, q, \lambda) - \nu\partial_q J(t, q, \lambda) + \kappa\nu^{\gamma+1}\lambda \} = 0, \quad (5.2.14)$$

Then, notice from (5.2.10) that $h(T, q, \lambda) = \pi\lambda q^2$ since $H(T, x, S, q, \lambda, N) = x + qS - \pi\lambda q^2$ and that $h(t, 0, \lambda) = 0$ since $H(t, x, S, 0, \lambda, N) = x$. Therefore, from the fact that ν^* is a global minimum point in equation (5.2.14) combined with the terminal and boundary conditions $J(T, q, \lambda) = \pi\lambda q^2$ and $J(t, 0, \lambda) = 0$, we can apply classical verification theorems such as Theorem 5.3 and Corollary 5.4 of Sennwald (2007) stating that $(\nu_t^*)_t$ is an optimal strategy and J is the value function h . $\square$

From Proposition 5.1 with $h = J$, we have that the first order condition (5.2.12) implies the following feedback form for the optimal trading rate ν^* with $(t, q, \lambda) \in \mathcal{S}$

$$\nu^*(t, q, \lambda) = \left(\frac{\partial_q h(t, q, \lambda) - \eta \partial_\lambda h(t, q, \lambda)}{(\gamma + 1)\lambda\kappa} \right)^{1/\gamma}, \quad (5.2.15)$$

which is well defined (and admissible for $\gamma = 1/2$) since $\lambda \geq \lambda_{\min} > 0$ and since h is $C^1(\mathcal{S})$. Hence, among all the uniformly bounded, non-negative and $\mathbb{F}$ -predictable admissible solutions for the trading strategy, the optimal one $(\nu_t^*)_t$ is deterministic as it does not depend on S , W nor N^ν . Note from the definition of the last trading time τ in (5.2.1), that we always have $\nu^*(t, q, \lambda) = 0$ for $t > \tau$. Upon substitution back in equation (5.2.11), we find for all $(t, q, \lambda) \in \mathcal{S}$

$$\partial_t h(t, q, \lambda) + \phi q^2 - \frac{(\partial_q h(t, q, \lambda) - \eta \partial_\lambda h(t, q, \lambda))^{\frac{\gamma+1}{\gamma}}}{(\lambda\kappa(\gamma+1))^{1/\gamma}} \left(\frac{\gamma}{\gamma+1} \right) - \alpha\lambda \partial_\lambda h(t, q, \lambda) = 0. \quad (5.2.16)$$

with terminal condition $h(T, q, \lambda) = \pi\lambda q^2$ and boundary condition $h(t, 0, \lambda) = 0$. This boundary condition is implied by the non-negativity of $(\nu_t^*)_t$ and by the fact that when the remaining inventory hits $q = 0$, the agent stops acquiring shares as of that stopping time and there is no penalty (she thus simply walks away with x in cash). Therefore, the first term of H in (5.2.10) is the accumulated cash of the strategy, the second is the marked-to-market book value of the remaining inventory and h is an extra penalty due to price impact stemming from optimally

liquidating the rest of the shares. Note that when λ_t^ν tends to 0, the agent can liquidate all her remaining shares at time t without risking to have an adverse impact on price, which is optimal, and we thus find the additional boundary condition $\lim_{\lambda \rightarrow 0} h(t, q, \lambda) = 0$ for the variable λ . However, specifying $\lambda_0 > 0$ prevents this situation from happening with the mean-reverting intensity (5.2.4), even if this boundary condition is useful numerically. Taking the square-root law $\gamma = 1/2$, we find the following admissible optimal trading rate for all $(t, q, \lambda) \in \mathcal{S}$

$$\nu^*(t, q, \lambda) = \frac{4}{(3\lambda\kappa)^2} (\partial_q h(t, q, \lambda) - \eta \partial_\lambda h(t, q, \lambda))^2$$

and the value function

$$\partial_t h(t, q, \lambda) + \phi q^2 - \frac{4}{27(\lambda\kappa)^2} (\partial_q h(t, q, \lambda) - \eta \partial_\lambda h(t, q, \lambda))^3 - \alpha \lambda \partial_\lambda h(t, q, \lambda) = 0. \quad (5.2.17)$$

Indeed, this choice of the square-root law $\gamma = 1/2$ (as empirically motivated in the literature for temporary price impact) ensures that the optimal trading rate is non-negative. The PDE (5.2.16) cannot be solved analytically but is solved easily using a finite difference scheme. This scheme however requires that λ_0 is sufficiently large to ensure numerical convergence ($\lambda_0 > 0.1$). Since the optimal rate of trading is deterministic, the exact optimal strategy $(\nu_t^*)_t$ and Q^{ν^*} can be computed at the inception $t = 0$ and are depicted in the following Figures 5.2.1, 5.2.2, 5.2.3 for various sets of parameters.

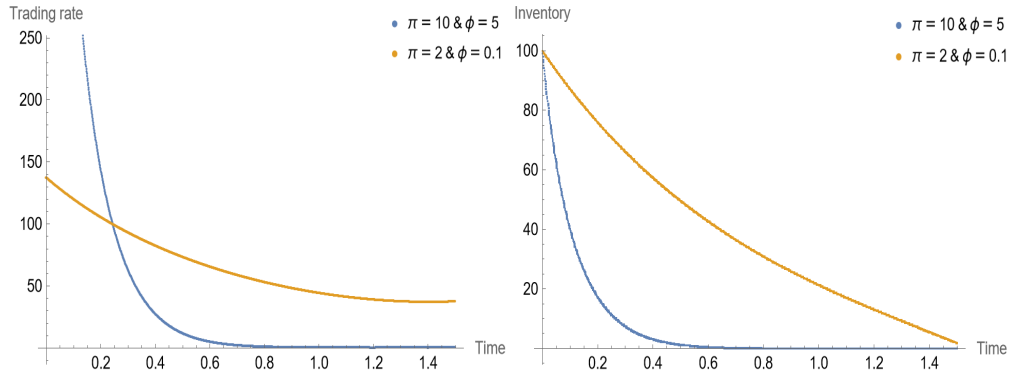


Figure 5.2.1: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact with $\pi = 10$ and $\phi = 5$ (blue line) compared to $\pi = 2$ and $\phi = 0.1$ (orange line) with $\kappa = 0.1$, $\eta = 0.002$, $\alpha = 0.5$, $\lambda_0 = 0.5$, $T = 1.5$ and $Q = 100$.

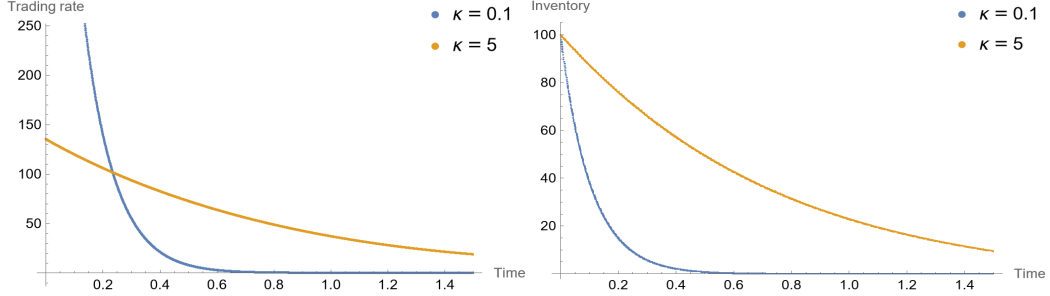


Figure 5.2.2: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact for different values of κ with $\pi = 10$, $\phi = 5$, $\eta = 0.001$, $\alpha = 0.5$, $\lambda_0 = 0.5$, $T = 1.5$ and $Q = 100$.

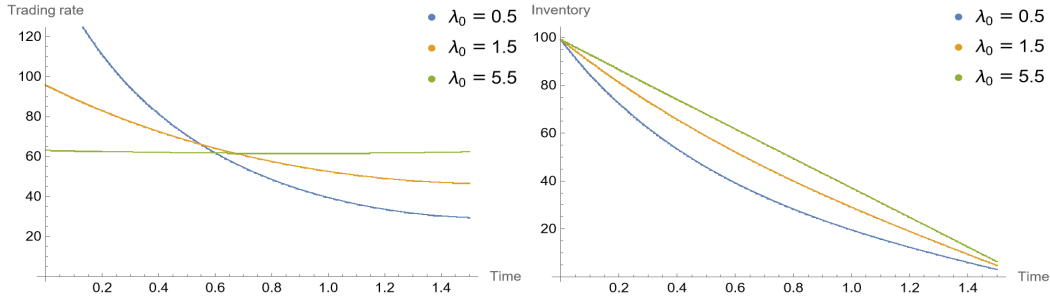


Figure 5.2.3: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact for various levels of λ_0 with $\pi = 10$, $\phi = 2$, $\eta = 0.001$, $\alpha = 0.5$, $\kappa = 0.1$, $T = 1.5$ and $Q = 100$.

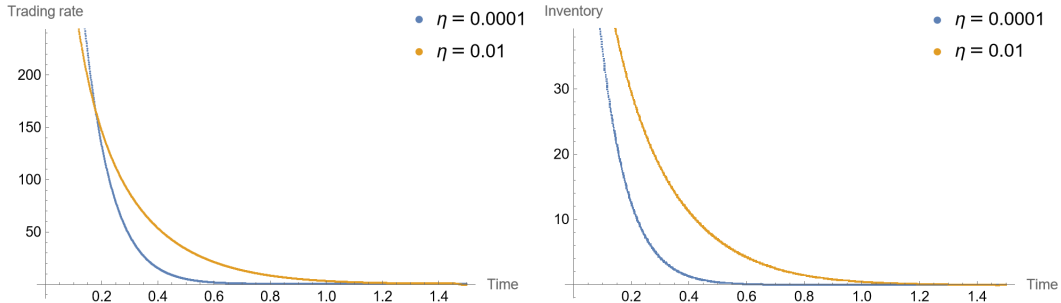


Figure 5.2.4: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary price impact for different values η with $\pi = 10$, $\phi = 5$, $\kappa = 0.1$, $\alpha = 0.5$, $\lambda_0 = 1.5$, $T = 1.5$, $Q = 100$.

If π and ϕ are sufficiently large (large terminal and running penalty), we see on the first Figure 5.2.1 that the agent has an incentive to liquidate her shares as soon as possible and we thus find in this case that $Q_T^{\nu^*} = 0$ (e.g. $\pi = 10$, $\phi = 5$). When π and ϕ are small, the agent has instead an incentive to liquidate her shares much slower and to keep a remaining quantity at maturity. Moreover, we observe on Figure 5.2.2 that when the temporary penalty κ incurred at each trade increases, the optimal agent has an incentive to liquidate less shares so as to mitigate the negative impact on the transaction price and to face instead the terminal penalty, which becomes small relative to this large temporary impact κ . Moreover, when adjusting the initial level λ_0 of the intensity (5.2.4), we observe significant variations in the trend of the trading rate even if it does not impact the value of the terminal inventory $Q_T^{\nu^*}$. Finally, when η is smaller and for large values of ϕ , we see that the remaining inventory decreases much faster at the beginning of the trading horizon. Indeed, since the (permanent) impact on the intensity λ^ν is much smaller, the agent can trade a lot at the beginning while keeping the intensity rather low (which is in accordance with the large urgency to trade ϕ). The impact of η depends however on the value of ϕ : when $\phi = 0$, we observe the opposite behavior (see the geometric intensity case 5.2.1.2 below with no running penalty) since in this case the agent has no urgency to trade and therefore no incentive to liquidate a lot at the beginning of the trading horizon even when η is low.

5.2.1.2 Temporary impact with geometric intensity

We now consider that the intensity λ^ν of the jump process N^ν follows instead the geometric dynamic

$$d\lambda_t^\nu = \eta \lambda_t^\nu \nu_t dt \quad \text{such that} \quad \lambda_t^\nu = \lambda_0^\nu \exp \left(\int_0^t \eta \nu_s ds \right). \quad (5.2.18)$$

We need to impose $\eta > 0$ so that a higher volume sold ν leads to an increase of the occurrence probability of a negative price impact (even if $\eta \leq 0$ could make sense empirically in some economic scenarios, see the above footnote¹). We also assume no more running penalty for detaining stocks throughout the trading horizon ($\phi = 0$). The agent is then willing to optimize the following performance criterion

$$\begin{aligned} H(t, x, S, q, \lambda) &= \sup_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t, x, S, q, \lambda} [X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu)] \\ &= x + qS - \inf_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t, q, \lambda} \left[\int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds + \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 \right] \\ &= x + qS - \inf_{(\nu_t)_t \in \mathcal{A}} h^\nu(t, q, \lambda) = x + qS - h(t, q, \lambda), \end{aligned} \quad (5.2.19)$$

with again $h(T, q, \lambda) = \pi \lambda q^2$, $h(t, 0, \lambda) = 0$ and $\lim_{\lambda \rightarrow 0} h(t, q, \lambda) = 0$ (since when the intensity is null, the agent can liquidate in an optimal way all her remaining

shares at zero cost). Again, specifying $\lambda_0 > 0$ in (5.2.18) prevents this situation from happening even if the limit $\lambda \rightarrow 0$ provides a boundary condition with respect to the variable λ necessary for solving numerically the HJB equation. Proposition 5.1 is still valid when replacing $(\eta\nu - \alpha\lambda)$ by $\eta\lambda\nu$. We therefore have that the following HJB equation satisfied by h for all $(t, q, \lambda) \in \mathcal{S}$ is a sufficient condition for the optimality of (5.2.19) :

$$\partial_t h(t, q, \lambda) + \min_{\nu \in \Gamma} \{ \eta\lambda\nu \partial_\lambda h(t, q, \lambda) - \nu \partial_q h(t, q, \lambda) + \lambda \kappa \nu^{\gamma+1} \} = 0 \quad (5.2.20)$$

with $h(T, q, \lambda) = \lambda \pi q^2$ and $h(t, 0, \lambda) = 0$. Taking $h(t, q, \lambda) = \lambda h(t, q)$, equation (5.2.20) becomes for all $(t, q) \in [0, T) \times (0, Q]$:

$$\partial_t h(t, q) + \min_{\nu \in \Gamma} \{ \eta \nu h(t, q) - \nu \partial_q h(t, q) + \kappa \nu^{\gamma+1} \} = 0, \quad (5.2.21)$$

and hence the optimization problem becomes independent of the intensity λ . This leads to

$$\nu^*(t, q) = \left(\frac{\partial_q h(t, q) - \eta h(t, q)}{(\gamma + 1)\kappa} \right)^{1/\gamma}, \quad (5.2.22)$$

and finally to

$$\partial_t h(t, q) - \frac{(\partial_q h(t, q) - \eta h(t, q))^{\frac{\gamma+1}{\gamma}}}{((\gamma + 1)\kappa)^{1/\gamma}} \left(\frac{\gamma}{\gamma + 1} \right) = 0,$$

for all $(t, q) \in [0, T) \times (0, Q]$ and with $h(T, q) = \pi q^2$ and $h(t, 0) = 0$. Hence, if we do not consider a running penalty anymore ($\phi = 0$) with the geometric intensity (5.2.18), the optimal rate $(\nu_t^*)_t$ does not depend on λ and is again purely deterministic (as it does not depend on W nor N^ν). We then find the following optimal strategies for various sets of parameters using the square-root law $\gamma = 1/2$.

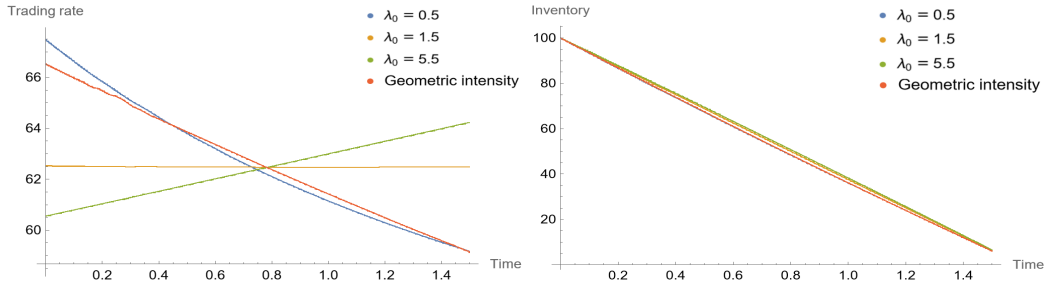


Figure 5.2.5: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact with the mean-reverting intensity (5.2.4) for various λ_0 (and $\phi = 0$), compared with the geometric intensity model (5.2.18) in red. Both models use $\pi = 8, \kappa = 0.8, \eta = 0.0025, \alpha = 0.05, Q = 100$ and $T = 1.5$.

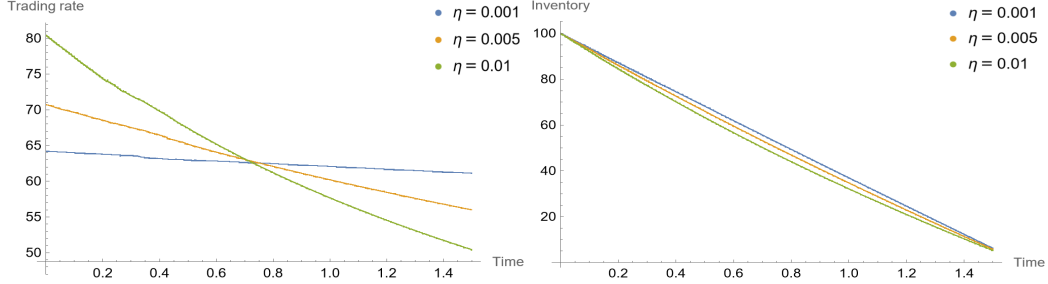


Figure 5.2.6: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact with geometric intensity (5.2.18) for various values of η with $\pi = 8$, $\kappa = 0.8$, $Q = 100$ and $T = 1.5$.

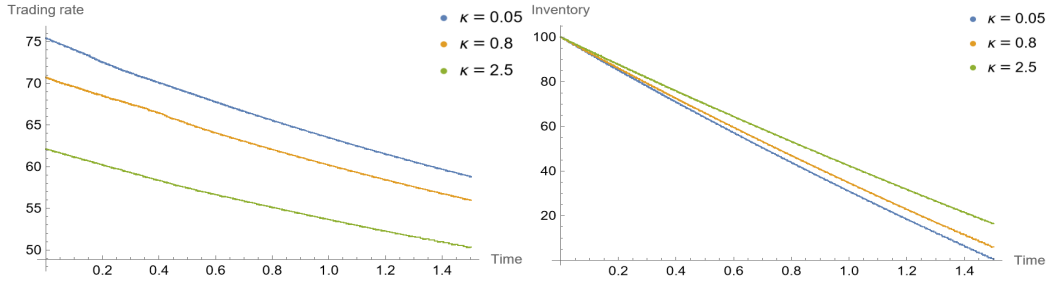


Figure 5.2.7: Optimal trading rate ν^* and inventory Q^{ν^*} from a temporary (direct) price impact with geometric intensity (5.2.18) for various values of κ with $\pi = 8$, $\eta = 0.005$, $Q = 100$ and $T = 1.5$.

We see in Figure 5.2.5 that the mean-reverting intensity model (5.2.4) from the previous Section 5.2.1.1 with $\lambda_0 = 0.5$ and the geometric intensity model (5.2.18) from this section lead to similar results. However, the convexity of the strategy is more pronounced when taking a mean-reverting intensity λ^ν in the optimization problem, while the slope of the trading rate tends to remain constant for the geometric intensity. For various values of λ_0 in the mean-reverting intensity (5.2.4), we see that the slope of $(\nu_t^*)_t$ with respect to time can change drastically leading to different kinds of strategies, while for the geometric intensity the trading rate is always decreasing with time. As already mentioned, since we consider no running penalty ($\phi = 0$), a higher η leads to a faster decrease of the shares to liquidate. Finally, we have logically that a higher temporary impact κ leads to a lesser trading rate throughout the trading horizon since the agent prefers in this case to face the terminal penalty and to avoid these higher negative impacts κ .

This temporary setting (5.2.6) with Dirac propagator G and mean-reverting intensity λ^ν appears as the most natural/consistent with the observations of Section 5.1 and Figure 5.1.2, as it allows to disentangle the temporary direct price impact that arises when MOs walk the LOB from the indirect permanent and transient impacts coming from the downward pressure in price due to changes in net order flow imbalance. However, we now propose in Section 5.2.2 and Section 5.2.3 other propagators that will allow to combine differently these temporary, transient and permanent (direct and indirect) effects.

5.2.2 Permanent price impact kernel

We consider in this section that the direct price impact ruled by the kernel G is permanent in the sense that a sell order permanently affects the price of the asset and hence the price at which all subsequent transactions are executed. This corresponds to setting $G \equiv 1$ in (5.2.3) which gives for $t \in [0, \tau]$:

$$dX_t^\nu = S_t^\nu \nu_t dt = \left(S_t - \int_0^t \kappa \nu_s^\gamma dN_s^\nu \right) \nu_t dt := (S_t - Z_t) \nu_t dt, \quad (5.2.23)$$

where $dS_t^\nu = dS_t - dZ_t = \sigma dW_t - \kappa \nu_t^\gamma dN_t^\nu$. Z_t is thus the cumulative negative price impact from the trades up to time t . Notice that the execution price $S^\nu = S - Z$ here makes sense per se when $G \equiv 1$, compared with the price (5.A.1) in the temporary case. Moreover, this setting allows to include both a permanent indirect impact (as before via λ^ν) as well as a permanent direct impact on prices via this propagator $G \equiv 1$. We again consider an agent willing to maximize the expected wealth at the last trading time τ :

$$\tilde{H}(t, x, S^\nu, q, \lambda) = \sup_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t, x, S^\nu, q, \lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_u^\nu)^2 du \right]. \quad (5.2.24)$$

We now consider that S^ν is a state variable (instead of S in the optimization (5.2.9) above) since the direct price impact in this section is permanent and hence, the price at which the agent can sell the Q_T^ν remaining shares at the end of the trading horizon is S_T^ν (including the cumulative negative price impacts) minus the terminal penalty. Finally, as mentioned above and in Appendix 5.A, this terminal penalty is a temporary impact as it only affects the last trade at T , and hence does not depend in a direct way on the previous negative jumps of N^ν (even if it still depends indirectly on the past trades $(\nu_s)_{0 \leq s \leq \tau}$ through the intensity λ^ν). This hence again leads to the same final penalty $\mathbb{E}_{t, x, S^\nu, q, \lambda} [Q_T^\nu (S_T^\nu - \pi \lambda_T^\nu Q_T^\nu)] = \mathbb{E}_{t, x, S^\nu, q, \lambda} [Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu)]$ as in Section 5.2.1, but with state variable S^ν instead of S . From equation (5.2.23), we can rewrite the wealth process at τ by

$$X_\tau^\nu = X_t^\nu + \int_t^\tau (S_s - Z_s) \nu_s ds.$$

Using integration by parts with $dQ_t^\nu = -\nu_t dt$, $dS_t = \sigma dW_t$ and $dZ_t = \kappa \nu_t^\gamma dN_t^\nu$, we have

$$X_\tau^\nu = X_t^\nu - S_\tau Q_\tau^\nu + S_t Q_t^\nu + Z_\tau Q_\tau^\nu - Z_t Q_t^\nu + \int_t^\tau \sigma Q_s^\nu dW_s - \int_t^\tau \kappa Q_s^\nu \nu_s^\gamma dN_s^\nu,$$

which leads to the *proceeds* R_τ^ν of the strategy at the stopping time τ (after which no more trading occurs),

$$\begin{aligned} R_\tau^\nu &:= X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu) \\ &= X_t^\nu + S_t Q_t^\nu - Z_t Q_t^\nu + \int_t^\tau \sigma Q_s^\nu dW_s - \int_t^\tau \kappa Q_s^\nu \nu_s^\gamma dN_s^\nu - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2. \end{aligned}$$

Again from the boundedness of Q^ν , the non-negativity of $Q^\nu \nu^\gamma$ and the $\mathbb{F}$ -predictability of $(\nu_t)_t$, we find

$$\mathbb{E}_{t,x,S^\nu,q,\lambda} [X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu)] = x + qS^\nu + \mathbb{E}_{t,q,\lambda} \left[- \int_t^\tau \kappa Q_s^\nu \nu_s^\gamma \lambda_s^\nu ds - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 \right].$$

We can then rewrite the value function (5.2.24) for $t \in [0, \tau]$ in the following way

$$\begin{aligned} \tilde{H}(t, x, S^\nu, q, \lambda) &= \sup_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t,x,S^\nu,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + qS^\nu - \inf_{(\nu_t)_t \in \mathcal{A}} \mathbb{E}_{t,q,\lambda} \left[\int_t^\tau \kappa Q_s^\nu \nu_s^\gamma \lambda_s^\nu ds + \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 + \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + qS^\nu - \inf_{(\nu_t)_t \in \mathcal{A}} \tilde{h}^\nu(t, q, \lambda) = x + qS^\nu - \tilde{h}(t, q, \lambda), \end{aligned} \quad (5.2.25)$$

where $\tilde{h}(t, q, \lambda)$ is again the cumulative penalty from optimally liquidating the rest of the shares, which is also bounded and non-negative in $[0, \tau]$ (and thus for all $(t, q, \lambda) \in \bar{\mathcal{S}}$). However, no regularity/continuity results exist for value functions of the form (5.2.25). From the boundedness of $\tilde{h}$ in $\bar{\mathcal{S}}$, we then use the theory of viscosity solutions instead, see Touzi (2012), Fleming and Soner (2006). As we now show, we need to impose $\gamma > 1$ in the permanent case and hence, the square-root law $\gamma = 1/2$ which holds in the temporary case cannot apply anymore with $G \equiv 1$.

5.2.2.1 Permanent impact with mean-reverting intensity

As in Section 5.2.1.1, we first assume that the intensity λ^ν of N^ν is mean-reverting and satisfies (5.2.4). Since $\tilde{h}(t, q, \lambda)$ is bounded on $\bar{\mathcal{S}}$, classical results from dynamic programming (see Bouchard and Touzi (2011) or Theorem 2.2.3 of Bouchard (2007)) imply that $\tilde{h}(t, q, \lambda)$ is the viscosity solution of the following HJB equation for all $(t, q, \lambda) \in S$:

$$\partial_t \tilde{h}(t, q, \lambda) + \phi q^2 + \min_{\nu \in \Gamma} \left\{ (\eta \nu - \alpha \lambda) \partial_\lambda \tilde{h}(t, q, \lambda) - \nu \partial_q \tilde{h}(t, q, \lambda) + \kappa q \nu^\gamma \lambda \right\} = 0, \quad (5.2.26)$$

subject to terminal and boundary conditions

$$\tilde{h}(T, q, \lambda) = \pi \lambda q^2, \quad \tilde{h}(t, 0, \lambda) = 0 \text{ and } \lim_{\lambda \rightarrow 0} \tilde{h}(t, q, \lambda) = 0.$$

Again the first boundary condition comes from the non-negativity of $(\nu_t)_t$ and the stopping time (5.2.1), leading to $\tilde{H}(t, x, S^\nu, 0, N, \lambda) = x$, and the second boundary condition is necessary numerically even if specifying $\lambda_0 > 0$ prevents from attaining this boundary. From (5.2.26) with $\gamma > 1$, the optimal trading rate is given explicitly in $\mathcal{S}$ by

$$\nu^*(t, q, \lambda) = \left(\frac{\partial_q \tilde{h}(t, q, \lambda) - \eta \partial_\lambda \tilde{h}(t, q, \lambda)}{\kappa q \lambda \gamma} \right)^{\frac{1}{\gamma-1}}, \quad (5.2.27)$$

which is again deterministic (no dependence on N^ν nor W) and highly similar to (5.2.15), except that we now have the remaining inventory q in the denominator which accounts for the permanent direct impact on the transaction price. We could indeed denote $\gamma = \gamma' + 1$ where γ' is the exponent in (5.2.15) (since $\gamma' > 0$ in the previous section while now we have $\gamma > 1$). We have an exploding trading rate at the time of market drop-out since $\lim_{q \rightarrow 0} \nu^*(t, q, \lambda) = +\infty$, which prevents any smooth solution to the HJB equation (with again $\nu^*(t, q, \lambda) = 0$ for $t > \tau$). However, from (5.2.2), this trading rate is still admissible for $\gamma = 3/2$ and hence, substituting (5.2.27) inside (5.2.26), we find that $\tilde{h}$ is a viscosity solution for all $(t, q, \lambda) \in \mathcal{S}$ and $\gamma > 1$ of the PDE

$$\partial_t \tilde{h}(t, q, \lambda) + \phi q^2 - \frac{\left(\partial_q \tilde{h}(t, q, \lambda) - \eta \partial_\lambda \tilde{h}(t, q, \lambda) \right)^{\frac{\gamma}{\gamma-1}}}{(\kappa \gamma q \lambda)^{1/(\gamma-1)}} \left(\frac{\gamma-1}{\gamma} \right) - \alpha \lambda \partial_\lambda \tilde{h}(t, q, \lambda) = 0.$$

with terminal condition $\tilde{h}(T, q, \lambda) = \pi \lambda q^2$ and boundary conditions $\tilde{h}(t, 0, \lambda) = 0$, $\lim_{\lambda \rightarrow 0} \tilde{h}(t, q, \lambda) = 0$. Choosing $\gamma = 3/2$ ensures the non-negativity of the optimal control since $\nu^*(t, q, \lambda) = \frac{4}{9} \left(\frac{\partial_q \tilde{h}(t, q, \lambda) - \eta \partial_\lambda \tilde{h}(t, q, \lambda)}{\kappa q \lambda} \right)^2$ in $\mathcal{S}$, leading to

$$\partial_t \tilde{h}(t, q, \lambda) + \phi q^2 - \frac{4}{27(\kappa q \lambda)^2} \left(\partial_q \tilde{h}(t, q, \lambda) - \eta \partial_\lambda \tilde{h}(t, q, \lambda) \right)^3 - \alpha \lambda \partial_\lambda \tilde{h}(t, q, \lambda) = 0.$$

This PDE is seemingly identical to the HJB equation in the temporary setting with square-root law (5.2.17), except again that the quantity q is now in the denominator $(\kappa q \lambda)^2$. Choosing $\gamma = 3/2$ is indeed consistent with the square root law $\gamma' = 1/2$ of the previous temporary/instantaneous price impact setting (Section 5.2.1), since we now take $G \equiv 1$ and hence consider instead the total/aggregated (direct) price impact of each MO. This choice leads to the following figures of optimal strategies for various parameters. Another reasonable assumption would be to take γ close to 1 and hence consider a linear permanent (direct) impact whenever L^ν jumps, so as to be consistent with the linear permanent impact function investigated in Almgren et al. (2005).

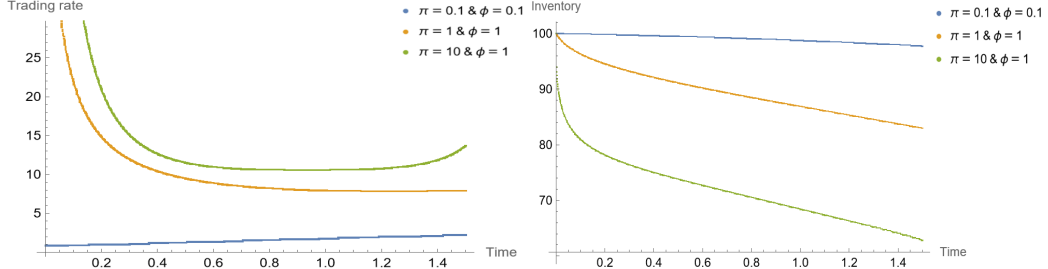


Figure 5.2.8: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact with mean-reverting intensity and $\gamma = 3/2$ for various values of π and ϕ with $\kappa = 1$, $\eta = 0.1$, $\alpha = 0.5$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

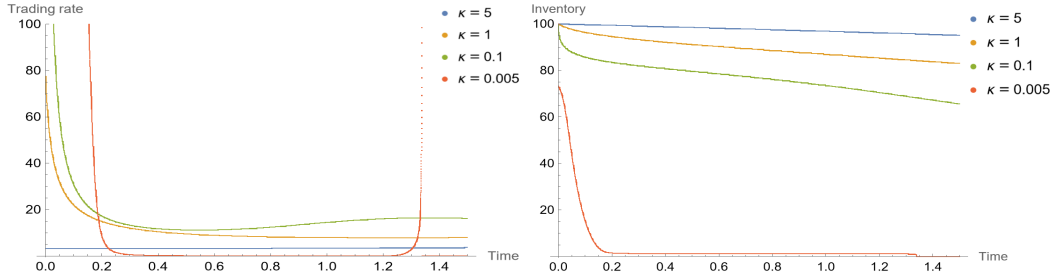


Figure 5.2.9: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact with mean-reverting intensity and $\gamma = 3/2$ for various values of κ with $\pi = 1$, $\phi = 1$, $\eta = 0.1$, $\alpha = 0.5$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

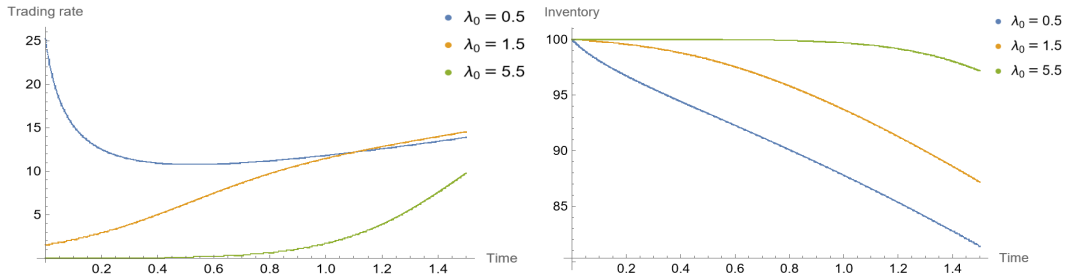


Figure 5.2.10: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact with mean-reverting intensity and $\gamma = 3/2$ for various values of λ_0 with $\pi = 1$, $\phi = 1$, $\kappa = 1$, $\eta = 0.1$, $\alpha = 1$, $Q = 100$ and $T = 1.5$.

Figures 5.2.8 and 5.2.9 again lead to financially consistent results when varying the penalty terms ϕ , π and κ . Moreover, when choosing really low values of κ , or high values of π and ϕ such as to liquidate all the shares before T , the optimal strategy is to liquidate most of the shares right at the beginning of the horizon. However, changing the intensity-related parameters λ_0 , α or η again strongly impacts the trend of the optimal strategy (see Fig. 5.2.10). Therefore, no clear pattern of optimal strategy arises with a permanent direct impact and mean-reverting intensity. Finally, we see that Fig. 5.2.9 (red line) confirms the exploding trading rate at the time of drop-out $q = 0$ when taking really low values κ to liquidate all the shares.

5.2.2.2 Permanent impact with geometric intensity

Secondly, we now assume for this permanent case $G \equiv 1$ no more running penalty ($\phi = 0$) and a geometric intensity linear in ν (as in Section 5.2.1.2) :

$$d\lambda_t^\nu = \eta \lambda_t^\nu \nu_t dt \quad \text{such that} \quad \lambda_t^\nu = \lambda_0 \exp \left(\int_0^t \eta \nu_s ds \right).$$

We again find in this geometric case that $\tilde{h}(t, q, \lambda)$ is a viscosity solution for all $(t, q, \lambda) \in \mathcal{S}$ of the HJB PDE

$$\partial_t \tilde{h}(t, q, \lambda) + \min_{\nu \in \Gamma} \left\{ \eta \nu \lambda \partial_\lambda \tilde{h}(t, q, \lambda) - \nu \partial_q \tilde{h}(t, q, \lambda) + \kappa q \nu^\gamma \lambda \right\} = 0,$$

with $\tilde{h}(T, q, \lambda) = \lambda \pi q^2$, $\tilde{h}(t, 0, \lambda) = 0$. Taking $\tilde{h}(t, q, \lambda) = \lambda \tilde{h}(t, q)$ leads to

$$\partial_t \tilde{h}(t, q) + \min_{\nu \in \Gamma} \left\{ \eta \nu \tilde{h}(t, q) - \nu \partial_q \tilde{h}(t, q) + \kappa q \nu^\gamma \right\} = 0, \quad (5.2.28)$$

with $\tilde{h}(T, q) = \pi q^2$ and $\tilde{h}(t, 0) = 0$. From (5.2.28) with $\gamma > 1$, the optimal trading rate satisfies for all $(t, q) \in [0, T] \times (0, Q]$:

$$\nu^*(t, q) = \left(\frac{\partial_q \tilde{h}(t, q) - \eta \tilde{h}(t, q)}{\kappa q \gamma} \right)^{\frac{1}{\gamma-1}}, \quad (5.2.29)$$

so that the optimal strategy $(\nu_t^*)_t$ is independent of λ (and again deterministic). Again, the obtained solution is highly similar to (5.2.22), except with the remaining inventory q in the denominator. This hence also leads to an exploding optimal trading rate $\nu^*(t, q, \lambda)$ at the time of market drop-out $q = 0$. Upon substitution back in the previous equation, we find that $\tilde{h}(t, q)$ is a viscosity solution for all $(t, q) \in [0, T] \times (0, Q]$ and $\gamma > 1$ of the following PDE

$$\partial_t \tilde{h}(t, q) - \frac{\left(\partial_q \tilde{h}(t, q) - \eta \tilde{h}(t, q) \right)^{\frac{\gamma}{\gamma-1}}}{(\kappa \gamma q)^{1/(\gamma-1)}} \left(\frac{\gamma-1}{\gamma} \right) = 0.$$

This model is numerically stable and gives the following optimal strategies starting at time $t = 0$ for various sets of parameters. Choosing $\gamma = 3/2$ is the most sensible choice so that the total market impact is consistent with the square-root law used in the temporary case, see Gatheral (2010). It also ensures the non-negativity and admissibility of $(\nu_t^*)_t$ in (5.2.29).

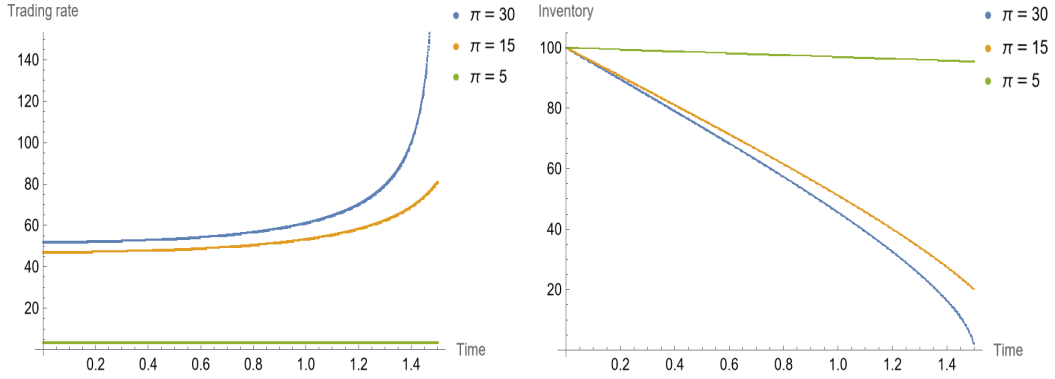


Figure 5.2.11: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact with geometric intensity and with $\gamma = 3/2$ for various values of π with $\kappa = 2$, $\eta = 0.01$, $Q = 100$ and $T = 1.5$.

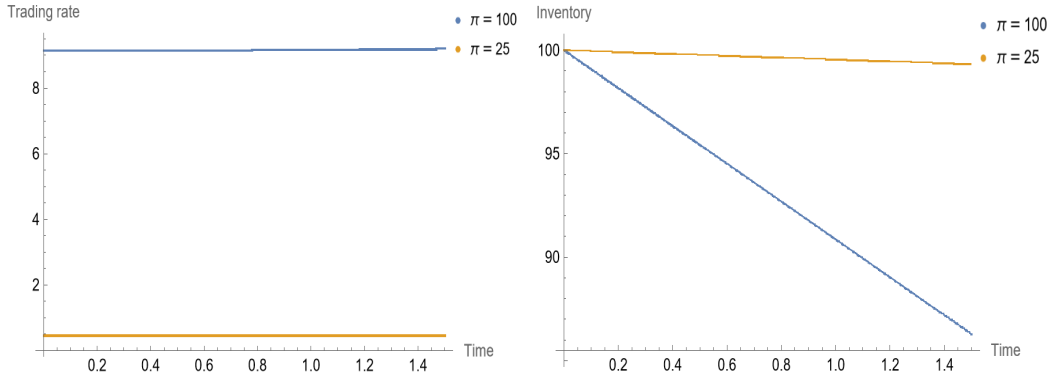


Figure 5.2.12: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact model with geometric intensity and $\gamma = 3/2$ for various values of π with $\kappa = 25$, $\eta = 0.01$, $Q = 100$ and $T = 1.5$.

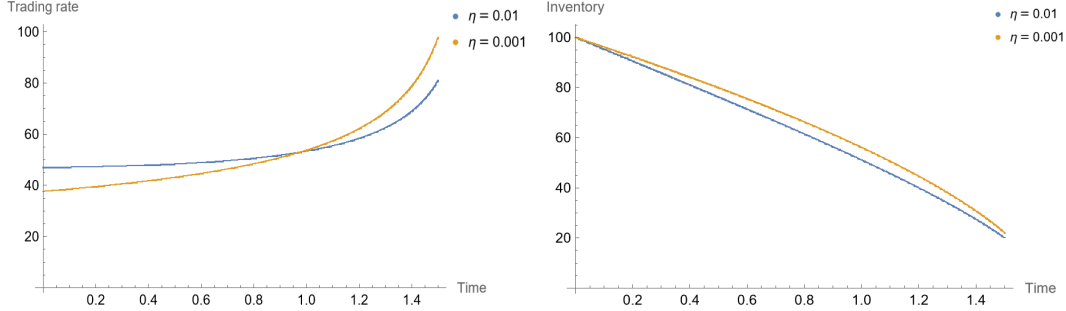


Figure 5.2.13: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a permanent (direct) price impact model with geometric intensity and $\gamma = 3/2$ for various values of η with $\kappa = 2$, $\pi = 15$, $Q = 100$ and $T = 1.5$.

We see in the first Figure 5.2.11 that when the terminal penalty π is sufficiently large and the permanent impact κ sufficiently small, the agent must liquidate all her shares before the trading horizon T so as to avoid the final penalty and to only incur the low permanent price impact κ during the trading horizon. However, to mitigate this (cumulative) permanent price impact on the whole trading horizon, the trading rate is increasing with time, which is the opposite strategy than in the temporary setting 5.2.1.2. Moreover, when κ is sufficiently low (or π large) so as to liquidate all the shares before T , the optimal strategy consists here in liquidating most of these shares at the end of the trading horizon. The explosion of the trading rate when the number of remaining shares tends to zero can indeed be seen in the blue curve with $\pi = 30$. We also observe that when π becomes small relative to κ (i.e. the ratio π/κ is sufficiently small), the agent prefers to incur the terminal penalty and to liquidate less shares during the trading horizon so as to mitigate the permanent impact. We can indeed see in Figure 5.2.11 (green line) a low constant trading rate coming from the very small terminal penalty. We also observe a similar constant trading rate in Figure 5.2.12 where the permanent impact κ is so large that the agent still prefers to incur the terminal penalty π , even if it is quite high. Of course, when the penalty π increases, the trading rate tends to increase as well so as to end up with a lower terminal inventory $Q_T^{\nu^*}$. Finally, we again have that the trend of the optimal strategy strongly depends on the intensity-related parameter (here η). Note that a mean-variance optimization setting can also be considered instead of the expected wealth maximization (5.2.24) studied so far, leading to similar results.

5.2.3 Transient impact function

In between the temporary case $G \equiv \delta$ and the permanent one $G \equiv 1$, we now propose two different ways of describing a transient direct price impact of the

agent's transactions.

5.2.3.1 First alternative

We consider at first a direct impact kernel of the form $G(t-s) = e^{-\beta(t-s)}$. Similarly as in Section 5.2.2.1, we can rewrite equation (5.2.23) by taking S and Z as state variables for $t \leq \tau$:

$$dX_t^\nu = \left(S_t - \kappa \int_0^t e^{-\beta(t-s)} \nu_s^\gamma dN_s^\nu \right) \nu_t dt = (S_t - Z_t) \nu_t dt, \quad (5.2.30)$$

where $dZ_t = -\beta Z_t dt + \kappa \nu_t^\gamma dN_t^\nu$ with $Z_0 = 0$ and where Z is again the cumulative negative impact from the trades on the transaction price. As in previous sections, we aim at maximizing the expected wealth of the agent given the set (5.2.2) of admissible trading strategies $\mathcal{A}$,

$$H^*(t, x, S, q, \lambda, Z) = \sup_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t, x, S, q, \lambda, Z} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau (Q_s^\nu)^2 ds \right],$$

where $S^\nu = S - Z$ is again used to define the terminal penalty since at the end of the trading horizon, the price at which we can sell the Q_T^ν remaining shares is S_T^ν (minus the penalty $\pi \lambda_T^\nu Q_T^\nu$) due to the transient direct price impact from G . We find the following proceeds R_τ^ν using (5.2.30) and integration by parts

$$\begin{aligned} R_\tau^\nu &= X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^\nu) \\ &= X_t^\nu + S_t Q_t^\nu - Z_t Q_t^\nu + \int_t^\tau Q_s^\nu (\sigma dW_s + \beta Z_s ds - \kappa \nu_s^\gamma dN_s^\nu) - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 \end{aligned} \quad (5.2.31)$$

since $\int_t^\tau Z_s \nu_s ds = -Z_\tau Q_\tau^\nu + Z_t Q_t^\nu + \int_t^\tau Q_s^\nu (-\beta Z_s ds + \kappa \nu_s^\gamma dN_s^\nu)$ and $\int_t^\tau S_s \nu_s ds = -S_\tau Q_\tau^\nu + S_t Q_t^\nu + \int_t^\tau Q_s^\nu \sigma dW_s$. We again have that $\int_t^\tau Q_s^\nu \sigma dW_s$ is a martingale since Q^ν is bounded and that $\mathbb{E}_{t, q, \lambda, Z} \left[\int_t^\tau \kappa Q_s^\nu \nu_s^\gamma dN_s^\nu \right] = \mathbb{E}_{t, q, \lambda, Z} \left[\int_t^\tau \kappa Q_s^\nu \nu_s^\gamma \lambda_s^\nu ds \right]$, see Lu (2013). We can hence rewrite H^* as

$$\begin{aligned} H^*(t, x, S, q, \lambda, Z) &= x + S^\nu q + \sup_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t, q, \lambda, Z} \left[\int_t^\tau Q_s^\nu (\beta Z_s - \kappa \nu_s^\gamma \lambda_s^\nu) ds - \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 - \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + S^\nu q - \inf_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t, q, \lambda, Z} \left[\int_t^\tau Q_s^\nu (\kappa \nu_s^\gamma \lambda_s^\nu - \beta Z_s) ds + \pi \lambda_\tau^\nu (Q_\tau^\nu)^2 + \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + S^\nu q - \inf_{(\nu_t)_{t \in \mathcal{A}}} h^{*, \nu}(t, q, \lambda, Z) = x + S^\nu q - h^*(t, q, \lambda, Z). \end{aligned} \quad (5.2.32)$$

However, we cannot make use of the following HJB equation for $(t, q, \lambda, Z) \in \mathcal{S} \times \mathbb{R}^+$

$$\begin{aligned} \partial_t h^*(t, q, \lambda, Z) + \phi q^2 - \beta Z \partial_Z h^*(t, q, \lambda, Z) - q\beta Z + \min_{\nu \in \Gamma} \left\{ (\eta\nu - \alpha\lambda) \partial_\lambda h^*(t, q, \lambda, Z) \right. \\ \left. - \nu \partial_q h^*(t, q, \lambda, Z) + q\kappa\nu^\gamma \lambda + \lambda [h^*(t, q, \lambda, Z + k\nu^\gamma) - h^*(t, q, \lambda, Z)] \right\} = 0, \end{aligned} \quad (5.2.33)$$

with $h^*(T, q, \lambda, Z) = \pi\lambda q^2$, $h^*(t, 0, \lambda, Z) = 0$, $\lim_{\lambda \rightarrow 0} h^*(t, q, \lambda, Z) = 0$ and $\gamma > 1$. Indeed, the value function $h^{*,\nu}(t, q, \lambda, Z)$ is decreasing in Z and hence not convex in ν . ν^* is then not a global minimum point of the argument in the minimization part of (5.2.33). Moreover, no explicit solution is available for the optimal trading rate ν^* since $\kappa\nu^\gamma$ is now inside of $h^*(t, q, \lambda, \cdot)$ in (5.2.33). This makes the solving of (5.2.33) numerically very intensive and unstable when using classical numerical schemes such as finite differences or finite elements. The following Chapter 6 will propose deep learning algorithms for solving such implicit optimal control problems.

5.2.3.2 Second alternative

We now give some alternative (and more specific) assumptions on the transient direct price impact and on the terminal penalty allowing to obtain the optimal trading rate explicitly (and deterministically). In the wealth (5.2.30), we assume in a first step that the dynamic of the price impact Z is instead given by

$$dZ_t = -\beta Z_t dt + \kappa Z_t \nu_t^\gamma dN_t^\nu, \quad (5.2.34)$$

with $\beta \geq 0$ and using Itô's lemma for jump processes, we obtain with $Z_0 > 0$:

$$Z_t = Z_0 e^{-\beta t} \prod_{0 \leq s \leq t}^{\Delta N_s^\nu \neq 0} (1 + \kappa \nu_s^\gamma),$$

leading to an execution price of the form $S_t^\nu = S_t - Z_t$ (where Z_0 is the half bid-ask spread, assumed here to be non-zero). However, we cannot write anymore the execution price in the form (5.2.3) based on a propagator G . We further assume that the intensity λ^ν of the jump process N^ν follows the mean-reverting dynamic $d\lambda_t^\nu = (\eta\nu_t - \alpha\lambda_t^\nu)dt$ as in (5.2.4). We finally assume that the terminal penalty when $Q_T^\nu > 0$ is a given multiple π' of the total negative impact Z_T at time T . The penalized price $S_T^{\pi'}$ at trading maturity T is thus defined formally as in (5.A.2) by the random variable

$$S_T^{\pi'} := S_T^\nu + dL_T^{\pi'} / dt = S_T^\nu - \pi' Z_T Q_T^\nu \sum_{j=1}^{N_T^\nu} \delta(T - t_j), \quad (5.2.35)$$

where $L_T^{\pi'} := -\pi' \int_0^T Q_s^\nu Z_{s-} dN_s^\nu$ with N^ν defined again as in (5.2.4). We see in (5.2.35) that π is replaced by $\pi' Z_T$ and it makes sense to impose $\pi' > 0$ since

we expect additional penalties that the agent must incur for the execution of the trade at the terminal date on top of the cumulative (transient) negative impact Z_T . Hence, we find as terminal penalty

$$\begin{aligned}\mathbb{E}_{t,x,S,q,\lambda,Z} \left[Q_T^\nu S_T^{\pi'} \right] &= \mathbb{E}_{t,x,S,q,\lambda,Z} [Q_T^\nu (S_T^\nu - \pi' Z_T \lambda_T^\nu Q_T^\nu)] \\ &= \mathbb{E}_{t,x,S,q,\lambda,Z} [Q_T^\nu (S_T^\nu - \pi' Z_T \lambda_T^\nu Q_T^\nu)].\end{aligned}\quad (5.2.36)$$

Following the exact same reasoning as for equation (5.2.31) on the dynamic (5.2.34) with $S^\nu = S - Z$, we then obtain the following optimization problem assuming $\phi = 0$ (no urgency to trade) :

$$\begin{aligned}H^*(t, x, S, q, \lambda, Z) &= \sup_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t,x,S,q,\lambda,Z} [X_\tau^\nu + Q_\tau^\nu (S_\tau^\nu - \pi' Z_\tau \lambda_\tau^\nu Q_\tau^\nu)] \\ &= x + S^\nu q - \inf_{(\nu_t)_{t \in \mathcal{A}}} \mathbb{E}_{t,q,\lambda,Z} \left[\int_t^\tau Q_s^\nu (\kappa Z_s \nu_s^\gamma \lambda_s^\nu - \beta Z_s) ds + \pi' Z_\tau \lambda_\tau^\nu (Q_\tau^\nu)^2 \right] \\ &= x + S^\nu q - \inf_{(\nu_t)_{t \in \mathcal{A}}} h^{*,\nu}(t, q, \lambda, Z) = x + S^\nu q - h^*(t, q, \lambda, Z),\end{aligned}\quad (5.2.37)$$

where $h^*(t, q, \lambda, Z)$ is again finite for all $(t, q, \lambda, Z) \in \bar{\mathcal{S}} \times \mathbb{R}^+$ (as well as non-decreasing in Z and non-negative whenever $\beta \leq \kappa \lambda_0 e^{-\alpha T} (\min_{s \in [t, \tau]} \nu_s^\gamma)$). We then again have that the value function $h^*(t, q, \lambda, Z)$ is a viscosity solution of the following HJB equation for $(t, q, \lambda, Z) \in \mathcal{S} \times \mathbb{R}^+$:

$$\begin{aligned}\partial_t h^*(t, q, \lambda, Z) - \beta Z \partial_Z h^*(t, q, \lambda, Z) - q \beta Z - \alpha \lambda \partial_\lambda h^* + \min_{\nu \in \Gamma} \left\{ \eta \nu \partial_\lambda h^* - \nu \partial_q h^* \right. \\ \left. + q \kappa Z \nu^\gamma \lambda + \lambda [h^*(t, q, \lambda, Z + \kappa Z \nu^\gamma) - h^*(t, q, \lambda, Z)] \right\} = 0,\end{aligned}\quad (5.2.38)$$

but now with $h^*(T, q, \lambda, Z) = \pi' \lambda Z q^2$, $h^*(t, 0, \lambda, Z) = 0$, $h^*(t, q, \lambda, 0) = 0$ and $\lim_{\lambda \rightarrow 0} h^*(t, q, \lambda, Z) = 0$. From the linearity of the value function (5.2.37) in Z , we then impose

$$h^*(t, q, \lambda, Z) = Z h_2^*(t, q, \lambda),$$

with $h_2^*(T, q, \lambda) = \pi' \lambda q^2$, $h_2^*(t, 0, \lambda) = 0$ and $\lim_{\lambda \rightarrow 0} h_2^*(t, q, \lambda) = 0$. Indeed, recall that in the previous Sections 5.2.1 and 5.2.2 on temporary and permanent (direct) impacts, the value functions h and $\tilde{h}$ could be seen as an extra penalty due to price impact stemming from optimally liquidating the rest of the shares. In this new transient setting where the terminal penalty is a multiple of the total negative impact, we therefore have that the value function h_2^* is still the extra penalty from optimally selling the remaining shares but expressed as a proportion of the total negative impact Z at that time. The value function $Z h_2^*(t, q, \lambda)$ is then again a viscosity solution for all $(t, q, \lambda, Z) \in \mathcal{S} \times \mathbb{R}^+$ of

$$\begin{aligned}Z \partial_t h_2^*(t, q, \lambda) - \beta Z h_2^*(t, q, \lambda) - q \beta Z - \alpha \lambda Z \partial_\lambda h_2^*(t, q, \lambda) \\ + \min_{\nu \in \Gamma} \left\{ \eta \nu Z \partial_\lambda h_2^*(t, q, \lambda) - \nu Z \partial_q h_2^*(t, q, \lambda) + q \kappa Z \nu^\gamma \lambda + \lambda \kappa Z \nu^\gamma h_2^*(t, q, \lambda) \right\} = 0,\end{aligned}$$

and therefore,

$$\begin{aligned} \partial_t h_2^*(t, q, \lambda) - \beta h_2^*(t, q, \lambda) - q\beta - \alpha\lambda\partial_\lambda h_2^*(t, q, \lambda) \\ + \min_{\nu \in \Gamma} \left\{ \eta\nu\partial_\lambda h_2^*(t, q, \lambda) - \nu\partial_q h_2^*(t, q, \lambda) + \lambda\kappa\nu^\gamma (q + h_2^*(t, q, \lambda)) \right\} = 0, \end{aligned} \quad (5.2.39)$$

We observe that ν^* is a global minimum point of the minimization part of the PDE (5.2.39) in $\mathcal{S}$ provided that $\gamma > 1$. We then have the following representation of the optimal trading rate for all $(t, q, \lambda) \in \mathcal{S}$ and $\gamma > 1$:

$$\nu^*(t, q, \lambda) = \left(\frac{\partial_q h_2^*(t, \lambda, q) - \eta\partial_\lambda h_2^*(t, \lambda, q)}{\lambda\kappa\gamma (q + h_2^*(t, \lambda, q))} \right)^{\frac{1}{\gamma-1}}. \quad (5.2.40)$$

Again, we have that $\lambda \geq \lambda_{\min} > 0$ due to the intensity (5.2.4) above, that $\lim_{q \rightarrow 0} \nu^*(t, q, \lambda) = +\infty$ and that $\nu^*(t, q, \lambda) = 0$ for $t > \tau$. We finally obtain that h_2^* is a viscosity solution in $\mathcal{S}$ of the following explicit PDE when $\gamma > 1$:

$$\begin{aligned} \partial_t h_2^*(t, q, \lambda) - \beta(h_2^*(t, q, \lambda) + q) - \alpha\lambda\partial_\lambda h_2^*(t, q, \lambda) \\ - \frac{(\partial_q h_2^*(t, q, \lambda) - \eta\partial_\lambda h_2^*(t, q, \lambda))^{\frac{\gamma}{\gamma-1}}}{(\lambda\kappa\gamma(q + h_2^*(t, q, \lambda)))^{1/(\gamma-1)}} \left(\frac{\gamma-1}{\gamma} \right) = 0. \end{aligned}$$

Choosing $\gamma = 3/2$ as in Section 5.2.2 on permanent impact is consistent with the square-root temporary impact and ensures the non-negativity and the admissibility of the optimal trading rate since we obtain $\nu^*(t, q, \lambda) = \frac{4}{9} \left(\frac{\partial_q h_2^* - \eta\partial_\lambda h_2^*}{\lambda\kappa(q + h_2^*)} \right)^2$ and

$$\partial_t h_2^*(t, q, \lambda) - \beta(h_2^*(t, q, \lambda) + q) - \alpha\lambda\partial_\lambda h_2^*(t, q, \lambda) - \frac{4(\partial_q h_2^*(t, q, \lambda) - \eta\partial_\lambda h_2^*(t, q, \lambda))^3}{27(\lambda\kappa(q + h_2^*(t, q, \lambda)))^2} = 0.$$

This leads to the following plots of optimal strategies in the transient price impact model (5.2.34).

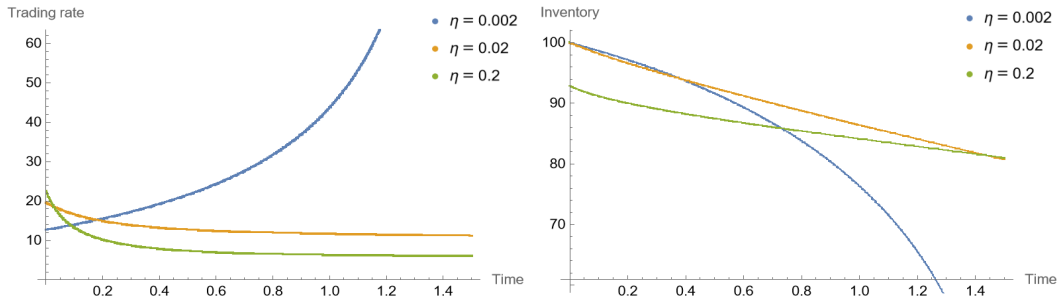


Figure 5.2.14: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a transient geometric (direct) price impact (5.2.34) with mean-reverting intensity and $\gamma = 3/2$ for various values of η with $\kappa = 0.006$, $\alpha = 0.5$, $\beta = 1$, $\pi' = 1000$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

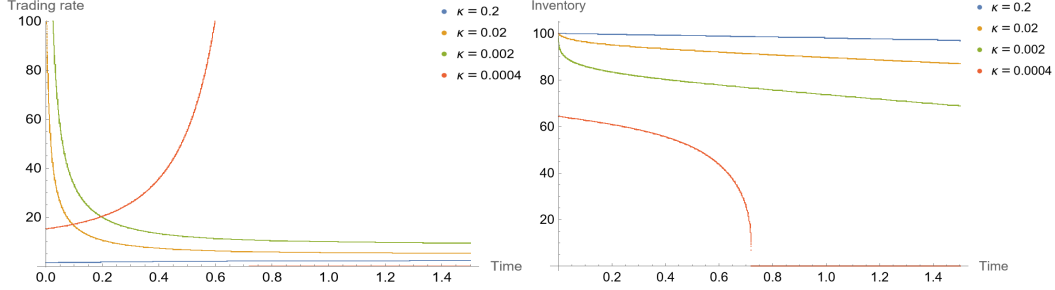


Figure 5.2.15: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a transient geometric (direct) price impact (5.2.34) with mean-reverting intensity and $\gamma = 3/2$ for various values of κ with $\alpha = 0.5$, $\eta = 0.1$, $\beta = 1$, $\pi' = 1000$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

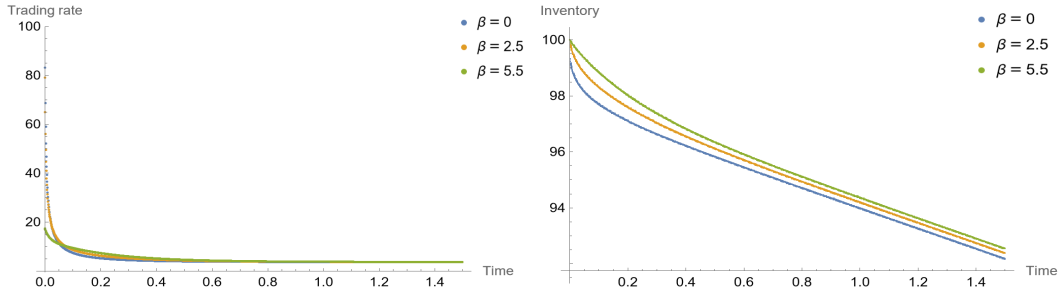


Figure 5.2.16: Optimal trading rate $(\nu_t^*)_t$ and inventory Q^{ν^*} from a transient geometric (direct) price impact (5.2.34) with mean-reverting intensity and $\gamma = 3/2$ for various values of β with $\alpha = 1$, $\eta = 0.5$, $\kappa = 0.05$, $\pi' = 100$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

Again, we see in Figure 5.2.14 that the intensity-related parameters strongly impact the trend of the optimal strategy. Moreover, when choosing a sufficiently small penalty κ such as to liquidate all the shares before T , we observe in Figure 5.2.15 that the optimal strategy consists in selling most of the shares right at the beginning of the trading horizon. This figure confirms the exploding behavior of the optimal trading rate at the time of market drop-out $q = 0$ when κ is really low (red line). Finally, the value of the drift β in (5.2.34) mostly influences the slope of the strategy at the beginning of the trading scheme.

5.3 Indirect optimal liquidation with limit and market orders

So far, we have considered that the liquidating agent could only post market (sell) orders. We now want to add the possibility for the seller to also post limit sell

orders (LOs) at a price $S_t + \delta_t$ that can be filled by MOs of buyers. The positive depth $(\delta_t)_t := (\delta_t)_{t \in [0, T]}$ at which the LO is posted is hence also a control process and we again aim at maximizing the expected wealth of the selling agent. Based on Guéant and Lehalle (2015), Guéant (2016) and Chapter 8.2 of Cartea et al. (2015), we expand the previous setting $dQ_t^\nu = -\nu_t dt$ of Section 5.2 by assuming instead

$$dQ_t^{\nu, \delta} = -\nu_t dt - \Delta dM_t^\delta, \quad (5.3.1)$$

where M^δ is a new (controlled) counting process corresponding to the number of buy MOs which lift the agent's LO, *i.e.* MOs which walk the sell side of the book to a price greater than or equal to $S_t + \delta_t$. Each filled LO is of size $\Delta > 0$. We consider that the number of arriving buy market orders (from other traders) follows a Poisson process M^0 with constant intensity Λ . The probability that the agent's LO at time t with depth δ_t will be lifted when a buy MO arrives (when M^0 jumps) is then assumed to be $\mathbb{P}(\delta_t) = e^{-\rho \delta_t}$, $\rho > 0$. Thus, the intensity of the counting process M^δ for filled limit orders is $\Lambda e^{-\rho \delta_t}$. The model (5.3.1) hence induces that the agent is continuously posting MOs at a rate $(\nu_t)_t$ and that her LOs at price $S_t + \delta_t$ are filled whenever M_t^δ jumps. For simplicity, we first assume as in Section 5.2.1 that G is purely temporary such that the execution price is given by $S_t^\nu = S_t - \kappa \nu_t^\gamma dN_t^\nu / dt$, with N^ν the counting process defined as previously with mean-reverting intensity (5.2.4). The wealth process is then obtained accordingly by

$$dX_t^{\nu, \delta} = (S_t + \delta_t) \Delta dM_t^\delta + (S_t^\nu - \xi) \nu_t dt,$$

where ξ is the half bid-ask spread (which is set to zero in the sequel, without loss of generality). We thus find for the wealth process at the last trading time τ on $[0, T]$:

$$X_\tau^{\nu, \delta} := X_t^{\nu, \delta} + \int_t^\tau (S_s + \delta_s) \Delta dM_s^\delta + \int_t^\tau S_s \nu_s ds - \int_t^\tau \kappa \nu_s^{\gamma+1} dN_s^\nu. \quad (5.3.2)$$

The optimization problem becomes in this setting

$$\begin{aligned} H(t, x, S, q, \lambda^\nu) &= \sup_{((\nu_t)_t, (\delta_t)_t) \in \mathcal{A}'} H^{\nu, \delta}(t, x, S, q, \lambda^\nu) \\ &= \sup_{((\nu_t)_t, (\delta_t)_t) \in \mathcal{A}'} \mathbb{E}_{t, x, S, q, \lambda^\nu} \left[X_\tau^{\nu, \delta} + Q_\tau^{\nu, \delta} \left(S_\tau - \pi \lambda_\tau^\nu Q_\tau^{\nu, \delta} \right) - \phi \int_t^\tau (Q_s^{\nu, \delta})^2 ds \right], \end{aligned}$$

where $\mathcal{A}'$ is defined as the set $\mathcal{A}$ in (5.2.2), enlarged with $\mathbb{F}$ -predictable, bounded from below depths $(\delta_t)_t$. The notation $\mathbb{E}_{t, x, S, q, \lambda^\nu}$ represents expectation conditional on $X_{t-}^{\nu, \delta} = x$, $S_t = S$, $Q_{t-}^{\nu, \delta} = q$ and $\lambda_t^\nu = \lambda^\nu$. If the agent has inventory remaining at the end of the trading horizon, she liquidates it using a MO for which she obtains worse prices than the midprice (terminal penalty). As in Section 5.2.1, we again need to include the intensity λ_T^ν in this terminal penalty and the terminal price at which she can sell the $Q_T^{\nu, \delta}$ shares is S_T (and not S_T^ν) since the direct impact of

MOs is here assumed to be temporary. Using integration by parts based on (5.3.1), we first find

$$\int_t^\tau S_s (\nu_s ds + \Delta dM_s^\delta) = -S_\tau Q_\tau^{\nu,\delta} + S_t Q_t^{\nu,\delta} + \int_t^\tau \sigma Q_s^{\nu,\delta} dW_s.$$

Hence, we directly obtain that

$$\begin{aligned} & \mathbb{E}_{t,x,S,q,\lambda^\nu} \left[X_\tau^{\nu,\delta} + Q_\tau^{\nu,\delta} (S_\tau - \pi \lambda_\tau^\nu Q_\tau^{\nu,\delta}) - \phi \int_t^\tau (Q_s^{\nu,\delta})^2 ds \right] \\ &= \mathbb{E}_{t,x,S,q,\lambda^\nu} \left[X_t^{\nu,\delta} + S_t Q_t^{\nu,\delta} + \int_t^\tau \sigma Q_s^{\nu,\delta} dW_s - \int_t^\tau \kappa \nu_s^{\gamma+1} dN_s^\nu + \int_t^\tau \delta_s \Delta dM_s^\delta \right. \\ & \quad \left. - \pi \lambda_\tau^\nu (Q_\tau^{\nu,\delta})^2 - \phi \int_t^\tau (Q_s^{\nu,\delta})^2 ds \right] \\ &= x + Sq - \mathbb{E}_{t,q,\lambda^\nu} \left[\int_t^\tau \kappa \nu_s^{\gamma+1} \lambda_s^\nu ds - \int_t^\tau \delta_s \Delta \Lambda e^{-\rho \delta s} ds + \pi \lambda_\tau^\nu (Q_\tau^{\nu,\delta})^2 + \phi \int_t^\tau (Q_s^{\nu,\delta})^2 ds \right] \\ &= x + Sq - h^{\nu,\delta}(t, q, \lambda^\nu), \end{aligned}$$

which leads to

$$H(t, x, S, q, \lambda^\nu) = x + qS - \inf_{((\nu_t)_t, (\delta_t)_t) \in \mathcal{A}'} h^{\nu,\delta}(t, q, \lambda^\nu) = x + qS - h(t, q, \lambda^\nu). \quad (5.3.3)$$

Given the admissible set $\mathcal{A}'$, $h(t, q, \lambda^\nu)$ is again bounded on $\bar{\mathcal{S}}$. Following Proposition 5.1 of Section 5.2.1 and from the verification theorem 3.4 of Guéant and Lehalle (2015) (see also Chapter 8.4 of Carlea et al. (2015)), we directly have that a sufficient condition for the optimality of (5.3.3) is for the C^1 -function $h(t, q, \lambda^\nu)$ to be a classical solution of the following HJB equation for all $(t, q, \lambda^\nu) \in \mathcal{S}$:

$$\begin{aligned} & \partial_t h + \phi q^2 - \alpha \lambda^\nu \partial_{\lambda^\nu} h + \min_{\nu \in \Gamma} \{ \eta \nu \partial_{\lambda^\nu} h - \nu \partial_q h + \kappa \nu^{\gamma+1} \lambda^\nu \} \\ & + \min_{\delta \in \Gamma'} \{ \Lambda e^{-\rho \delta} [h(t, q - \Delta, \lambda^\nu) - h(t, q, \lambda^\nu)] - \delta \Delta \Lambda e^{-\rho \delta} \} = 0, \end{aligned} \quad (5.3.4)$$

with terminal condition $h(T, q, \lambda^\nu) = \pi \lambda^\nu q^2$ and boundary conditions $h(t, 0, \lambda^\nu) = 0$, $\lim_{\lambda^\nu \rightarrow 0} h(t, q, \lambda^\nu) = 0$. Finally, from the first order condition in (5.3.4) with $\gamma > 0$, we find the exact same expression for ν^* as equation (5.2.15):

$$\nu^*(t, q, \lambda^\nu) = \left(\frac{\partial_q h(t, q, \lambda^\nu) - \eta \partial_{\lambda^\nu} h(t, q, \lambda^\nu)}{(\gamma + 1) \lambda^\nu \kappa} \right)^{1/\gamma}, \quad (5.3.5)$$

with $\nu^*(t, q, \lambda^\nu) = 0$ for $t > \tau$, and for the optimal depth δ^* we obtain

$$\delta^*(t, q, \lambda^\nu) = \frac{1}{\rho} + \frac{h(t, q - \Delta, \lambda^\nu) - h(t, q, \lambda^\nu)}{\Delta}.$$

Substituting these optimal strategies back into the HJB equation (5.3.4), we finally find that h must satisfy for all $(t, q, \lambda^\nu) \in \mathcal{S}$ and $\gamma > 0$ the explicit PDE

$$\begin{aligned} \partial_t h(t, q, \lambda^\nu) + \phi q^2 - \alpha \lambda^\nu \partial_{\lambda^\nu} h(t, q, \lambda^\nu) - \frac{(\partial_q h(t, q, \lambda^\nu) - \eta \partial_{\lambda^\nu} h(t, q, \lambda^\nu))^{\frac{\gamma+1}{\gamma}}}{(\lambda^\nu \kappa (\gamma+1))^{1/\gamma}} \left(\frac{\gamma}{\gamma+1} \right) \\ - \frac{\Lambda e^{-1} \Delta}{\rho} \exp \left(-\rho \left[\frac{h(t, q - \Delta, \lambda^\nu) - h(t, q, \lambda^\nu)}{\Delta} \right] \right) = 0, \end{aligned} \quad (5.3.6)$$

which can be solved numerically in a very efficient way using the method of lines (with a delayed PDE in q). Again, choosing the square root law $\gamma = 1/2$ appears to be the most consistent choice in the temporary case. As the remaining inventory q now depends on the counting process M^δ through (5.3.1), the optimal trading rate $\nu^*(t, q, \lambda^\nu)$ and the optimal depth $\delta^*(t, q, \lambda^\nu)$ in the LOB are not deterministic anymore. Therefore, we now have a more realistic price impact model where the liquidating agent can both use MOs and LOs. We then obtain in Figure 5.3.1 the following optimal strategies $(\nu_t^*)_t$ and $(\delta_t^*)_t$ for various sets of parameters.

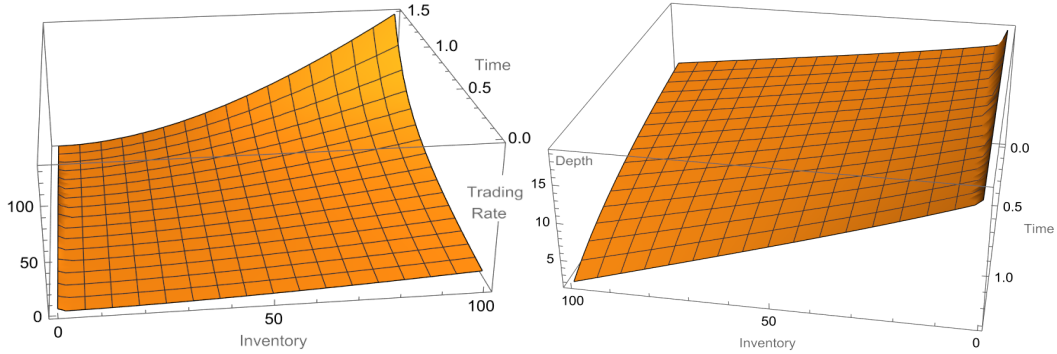


Figure 5.3.1: Optimal trading rate $(\nu_t^*)_t$ (left plot) and depth $(\delta_t^*)_t$ (right plot) in function of the remaining inventory q and time t at $\lambda^\nu = 2$ from a temporary (direct) price impact model allowing MOs and LOs with $\gamma = 1/2$, $\alpha = 0.5$, $\eta = 0.01$, $\kappa = 0.5$, $\pi = 0.05$, $\phi = 0$, $\Lambda = 1$, $\rho = 0.08$, $\Delta = 1$, $\lambda_0 = 0.5$, $Q = 100$ and $T = 1.5$.

We see on the first plot that the optimal trading rate $(\nu_t^*)_t$ is logically increasing with time (the closer the maturity, the more you need to trade so as to avoid the terminal penalty), and also increasing with the remaining inventory (the lower q , the less you need to trade so as to liquidate the remaining shares). We observe the opposite behavior for the optimal depth $(\delta_t^*)_t$, which is decreasing in time due to the agent becoming more averse to holding inventories as the trading horizon approaches, and also decreasing in inventory. Indeed, if the agent's inventory is

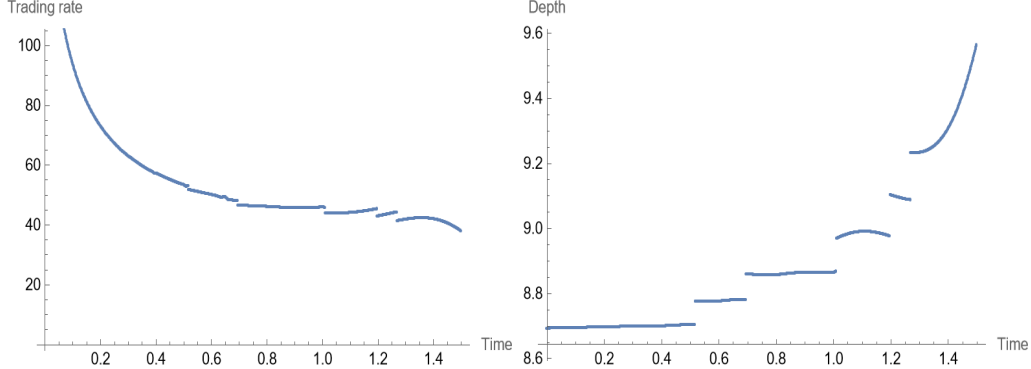


Figure 5.3.2: Optimal strategy of trading rate $(\nu_t^*)_t$ (left plot) and depth $(\delta_t^*)_t$ (right) from a temporary price impact model allowing MOs and LOs with $\gamma = 1/2$, $\alpha = 0.5$, $\eta = 0.01$, $\kappa = 0.5$, $\pi = 0.1$, $\phi = 0$, $\Lambda = 1$, $\rho = 0.08$, $\Delta = 1$, $\lambda_0 = 0.5$, $Q = 100$, $T = 1.5$.

large, she is willing to accept a lower premium δ for providing liquidity so as to increase the probability that her order is filled. At the same time, this ensures that she may complete the liquidation of the Q shares before T and hence avoids paying a terminal penalty. However, if inventories are low, the agent prefers setting a large depth δ because with low inventory, the terminal penalty she incurs when crossing the spread with MOs will be moderate. Figure 5.3.2 below confirms that the optimal trading rate and optimal depth both depend on the jumps of the counting process M^δ and are therefore not deterministic anymore.

This setting can be easily extended to a more general price impact model $S^\nu = S - Z$ with Z the cumulative price impact, by using instead a wealth process which depends on S^ν , *i.e.* :

$$dX_t^{\nu,\delta} = (S_t^\nu + \delta_t)\Delta dM_t^\delta + (S_t^\nu - \xi)\nu_t dt,$$

and by maximizing the expected wealth at τ (with S^ν in the terminal penalty),

$$\begin{aligned} & \mathbb{E}_{t,x,S,q,\lambda^\nu,Z} \left[X_\tau^{\nu,\delta} + Q_\tau^{\nu,\delta} \left(S_\tau^\nu - \pi \lambda_\tau^\nu Q_\tau^{\nu,\delta} \right) - \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + S^\nu q - \mathbb{E}_{t,q,\lambda^\nu,Z} \left[\int_t^\tau Q_s^{\nu,\delta} dZ_s - \int_t^\tau \delta_s \Delta \Lambda e^{-\rho \delta_s} ds + \pi \lambda_\tau^\nu \left(Q_\tau^{\nu,\delta} \right)^2 + \phi \int_t^\tau (Q_s^\nu)^2 ds \right] \\ &= x + S^\nu q - \tilde{h}^{\nu,\delta}(t, q, \lambda^\nu, Z). \end{aligned}$$

In the permanent case 5.2.2 with $G \equiv 1$, we find for example

$$\begin{aligned} \tilde{h}^{\nu,\delta}(t, q, \lambda^\nu, Z) \\ = \mathbb{E}_{t,q,\lambda^\nu} \left[\int_t^\tau \kappa Q_s^{\nu,\delta} \nu_s^\gamma \lambda_s^\nu ds - \int_t^\tau \delta_s \Delta \Lambda e^{-\rho \delta_s} ds + \pi \lambda_\tau^\nu \left(Q_\tau^{\nu,\delta} \right)^2 + \phi \int_t^\tau (Q_s^\nu)^2 ds \right]. \end{aligned}$$

Finally, as for the occurrence of price impact and LOs, we could have made the occurrence of MOs discrete instead of continuous. Indeed, continuous trading rates are obtained as limits of discrete trades in the high-frequency regime and those limiting results tend to produce continuous propagator models. This should be investigated in further research. Moreover, to obtain fully consistent and implementable strategies, the indirect price impact model introduced in this chapter should also be enhanced by including the possibility for the trader to take into account signals stemming from the market order-flow, see Fouque et al. (2022) or Cartea et al. (2018). Incorporating a short-term alpha component in the dynamic order execution schedule indeed enables the trader to leverage some knowledge and to execute more speculative trades. This is usually achieved by making the unaffected price S in equation (5.2.3) non-Martingale, and more precisely by adding a predictable zero-mean reverting process $(\alpha_t)_{0 \leq t \leq T}$ as drift in S , see *e.g.* Abi Jaber and Neuman (2022). In our setting, a more optimal approach would be to have the intensity λ^ν depend on the volume posted on both sides of the LOB. Indeed, the authors in Cartea et al. (2018) propose a new measure of volume imbalance in the LOB and demonstrate its accuracy as predictor of buying/selling pressures in the market. Given that the intensity λ^ν precisely models this pressure from net order-flow imbalance in our context, making it dependent on the volume imbalance measure of Cartea et al. (2018) should allow us to easily integrate signals into our execution strategies.

5.4 Conclusion

This paper introduces a new jump-dependent price impact model with propagator allowing to best reproduce the empirical direct and indirect effects of market orders on the transaction price, while including a stochastic liquidity factor in the form of a trade-driven intensity process. More precisely, the jump process is used in our setting to model the occurrence of the (temporary/permanent/transient) price impact arising from MOs since all trades do not lead to a direct negative impact on the transaction price, even if they systematically exert an (indirect) downward pressure on this price. This indirect downward price effect from order flow imbalance is here modeled by an increased mean-reverting intensity λ^ν of this jump process N^ν with negative jump size (which scale as a power-law function of the trading rate). Moreover, as the direct impact of market orders depends on the

propagator G in (5.2.3), we discuss for each choice of G the most relevant optimal liquidation problem faced by the agent, derive the related HJB equation and solve it numerically. In the temporary case $G \equiv \delta$, we find that a square-root impact provides the most consistent results and obtain a deterministic decreasing optimal trading strategy that can be set up at time $t = 0$. Moreover, when choosing a geometric intensity (5.2.18), we show that the dependence of this optimal strategy on the intensity λ^ν disappears. Secondly, in the permanent setting (5.2.23) with $G \equiv 1$ and exponent $\gamma = 3/2$, we again find a deterministic optimal trading strategy whose trend strongly varies according to the model parameters and which does not depend on λ^ν when considering again the geometric intensity (5.2.18). At last, in the transient case, alternative assumptions on the direct price impact and on the terminal penalty have to be assumed in order to obtain the optimal trading rate explicitly (and deterministically). In both the permanent and the transient direct settings, we obtain exploding trading rates at the time of market drop-out $q = 0$. Finally, these three scenarios based solely on MOs can be expanded by allowing the agent to post limit orders in the LOB at a depth δ , which is also a control process to be optimized. Optimal strategies of trading rates $(\nu_t^*)_t$ and depths $(\delta_t^*)_t$ are again obtained explicitly in this enhanced setting, leading to financially sound results. Our indirect price impact model with propagator therefore allows to offer sufficient tractability and flexibility so as to best reproduce the observed underlying dynamic of the LOB and its liquidity, thereby bridging the gap with high-frequency price models.

Subsequent papers should provide a comprehensive methodology for estimating the various model parameters based on real LOB data and determining which kernel G (and which exponent γ) allows to best fit the data (provided that high-frequency data of sufficient quality can be obtained). Secondly, we should incorporate a short-term alpha signal with discretized MOs' trading rates into our dynamic execution schedule in order to obtain more realistic and implementable liquidation strategies. Moreover, deep learning methods will be investigated and applied in Chapter 6 for solving high-dimensional HJB equations without explicit solutions, as obtained with equations (5.2.33) or (5.B.2). Finally, price manipulation with the no-dynamic-arbitrage properties of the introduced indirect price impact model should also be assessed for the various propagators G studied in this paper.

5.A Dirac's delta propagator

This appendix shows the equivalence in the temporary case between the execution price (5.2.3) with Dirac propagator $G(t-s) = \delta(t-s)$ and the wealth process introduced in (5.2.6). We can indeed write formally in this case the execution price

$$\begin{aligned} S_t^\nu &= S_t + \int_0^t G(t-s) dL_s^\nu = S_t - \kappa \int_0^t G(t-s) \nu_s^\gamma dN_s^\nu \\ &= S_t - \kappa \sum_{j=1}^{N_t^\nu} \nu_{t_j}^\gamma \delta(t-t_j) \\ &= S_t + dL_t^\nu/dt, \end{aligned} \quad (5.A.1)$$

where t_j ($j = 1, \dots, N_t^\nu$) are the jump times of N^ν over $[0, \tau]$. This indeed leads to the expression (5.2.6) of the wealth process $X_t^\nu = \int_0^t S_s \nu_s ds + \int_0^t \frac{dL_s^\nu}{ds} \nu_s ds = \int_0^t S_s \nu_s ds - \int_0^t \kappa \nu_s^{\gamma+1} dN_s^\nu$. Such process S^ν for the execution price (5.A.1) is sometimes called a "*Dirac process*" and has already been investigated in Kenyon and Green (2015) and Stein et al. (2022) as arising from the time derivative of a Poisson process (in a similar way as for impulse control problems). This form of the execution price S^ν is not interesting per se and cannot be used to define our model (just as a Brownian motion alone is not particularly useful, a Dirac process is only useful for what can be done with it). However, this price impact model (5.A.1) is justified by the fact that it is consistent with the definition of the wealth process (5.2.6), which is the only reasonable form for X^ν in the temporary case and which is the process of interest in the optimization program below (the state variables in the temporary setting are indeed S and X^ν , but never S^ν).

We also derive the terminal penalty $Q_T^\nu(S_T - \pi \lambda_T^\nu Q_T^\nu)$ used in the optimization problem (5.2.9). Consistently with the literature where the temporary price impact in terms of the trading rate $\kappa \nu_t^\gamma$ is replaced by a penalty πQ_T^ν at terminal time T , see Cartea and Jaimungal (2016), Cartea et al. (2015) and Donnelly (2022), we introduce our penalized price S_T^π at T as a sum of Dirac's delta according to the execution price (5.A.1) but with jump size $-\pi Q_T^\nu$ instead of $-\kappa \nu_t^\gamma$ in the following (formal) way

$$S_T^\pi := S_T + dL_T^\pi/dt = S_T - \pi Q_T^\nu \sum_{j=1}^{N_T^\nu} \delta(T-t_j), \quad (5.A.2)$$

where $L_T^\pi := -\pi \int_0^T Q_s^\nu dN_s^\nu$ is a new random variable for the terminal penalty at time T and with N^ν the counting process (5.2.4). Note that even when the direct price impact is not temporary but permanent or transient (see Sections 5.2.2 and 5.2.3 with other kernels G), the terminal penalty remains a temporary direct impact (with Dirac delta $\delta(T-t_j)$) since it only affects the price of the transaction at final

time T and not the other ones. With a penalized price of the form (5.A.2), we retrieve the terminal penalty

$$\begin{aligned}\mathbb{E}_{t,x,S,q,\lambda} [Q_T^\nu S_T^\pi] &= \mathbb{E}_{t,x,S,q,\lambda} [Q_T^\nu (S_T - \pi \lambda_T^\nu Q_T^\nu)] \\ &= \mathbb{E}_{t,x,S,q,\lambda} [Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu)] .\end{aligned}\quad (5.A.3)$$

Again the expression (5.A.2) cannot be used to define our model but is induced and justified by the chosen form of the terminal penalty in the optimization (5.2.9).

5.B Alternative risk-aversion penalty

As already explained in Section 5.2, the term $\int_t^\tau (Q_u^\nu)^2 du$ in (5.2.9) can be seen as a form of risk-aversion (urgency to trade) of the agent. We now propose another form of agent's risk aversion based on an expected shortfall penalty allowing to limit the loss on the total revenue of the liquidating agent at the end of the trading τ . For simplicity, we consider the temporary setting of Section 5.2.1.1 but the following results can be directly extended to other propagators G . We hence suppose the following performance criterion

$$H^\nu(t, x, S, q, \lambda) = \mathbb{E}_{t,x,S,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi [K' - X_\tau^\nu]^+ \right]. \quad (5.B.1)$$

We indeed penalize the agent only when the revenue X_τ^ν at the last trading time τ is below a threshold K' . This penalization term $-\phi [K' - x]^+$ is Lipschitz continuous and concave in x and it can then be shown that the bounded value function $h(t, x, S, q, \lambda) = \sup_{(\nu_t)_t \in \mathcal{A}} H^\nu(t, x, S, q, \lambda)$ must be a viscosity solution of the following HJB equation for all $(t, q, \lambda, x, S) \in \mathcal{S} \times \mathbb{R} \times \mathbb{R}$:

$$\begin{aligned}\partial_t h + \frac{1}{2} \sigma^2 \partial_S^2 h + \sup_{\nu \in \Gamma} \left\{ (\eta \nu - \alpha \lambda) \partial_\lambda h - \nu \partial_q h \right. \\ \left. + \nu S \partial_x h + \lambda [h(t, x - \kappa \nu^{\gamma+1}, S, q, \lambda) - h(t, x, S, q, \lambda)] \right\} = 0,\end{aligned}\quad (5.B.2)$$

with $h(T, x, S, q, \lambda) = x - \phi [K' - x]^+ + qS - \pi \lambda q^2$. This 5-dimensional HJB equation is however hard to solve numerically since we only have an implicit solution for $(\nu_t^*)_t$, see Chapter 6. We therefore consider instead the following performance criterion

$$H^\nu(t, x, S, q, \lambda) = \mathbb{E}_{t,x,S,q,\lambda} \left[X_\tau^\nu + Q_\tau^\nu (S_\tau - \pi \lambda_\tau^\nu Q_\tau^\nu) - \phi \int_t^\tau \nu_u [K - S_u]^+ du \right], \quad (5.B.3)$$

with again $h(t, x, S, q, \lambda) = \sup_{(\nu_t)_t \in \mathcal{A}} H^\nu(t, x, S, q, \lambda)$. Indeed, with this form of running penalty, we can show that if $(K' - X_\tau) < \int_t^\tau \nu_u (K - S_u) du$, then we have

$$(K' - X_\tau)^+ \leq \left(\int_t^\tau \nu_u (K - S_u) du \right)^+ \leq \int_t^\tau \nu_u (K - S_u)^+ du,$$

and thus that

$$\phi(K' - X_\tau)^+ \leq \phi \int_t^\tau \nu_u (K - S_u)^+ du,$$

which ensures that the penalty in (5.B.3) is at least larger than the shortfall penalty (5.B.1) and hence, that the optimization (5.B.3) also limits the loss on the revenue X_τ^ν at the final trading time τ . We therefore need to impose

$$\begin{aligned} K' - \int_t^\tau S_u \nu_u du + \int_t^\tau \kappa \nu_u^{\gamma+1} dN_u^\nu &< \int_t^\tau \nu_u (K - S_u) du \\ \Leftrightarrow K' &< \int_t^\tau K \nu_u du - \int_t^\tau \kappa \nu_u^{\gamma+1} dN_u^\nu, \end{aligned}$$

which is verified when assuming

$$K > \frac{K' + \kappa (\max_{s \in [t, \tau]} \nu_s^{\gamma+1}) (N_\tau^\nu - N_t^\nu)}{Q_t^\nu - Q_\tau^\nu}, \quad (5.B.4)$$

since $(\nu_t)_t \in \mathcal{A}$ is non-negative and bounded on $[0, \tau]$, $\gamma > 0$. The condition (5.B.4) is systematically satisfied empirically when choosing $K > K'/Q_t^\nu$ as the agent aims to liquidate a large number of shares in a short interval of time T . As in Section 5.2.1, we can rewrite the value function (5.B.3) by $h(t, x, S, q, \lambda) = x + qS + \tilde{h}(t, S, q, \lambda)$ and show that $\tilde{h}$ is a viscosity solution of the HJB PDE for all $(t, q, \lambda, S) \in \mathcal{S} \times \mathbb{R}$:

$$\partial_t \tilde{h} + \frac{1}{2} \sigma^2 \partial_{SS}^2 \tilde{h} + \sup_\nu \left\{ -\phi \nu [K - S]^+ + (\eta \nu - \alpha \lambda) \partial_\lambda \tilde{h} - \nu \partial_q \tilde{h} - \lambda \kappa \nu^{\gamma+1} \right\} = 0,$$

with terminal/boundary conditions $\tilde{h}(T, S, q, \lambda) = -\pi \lambda q^2$, $\tilde{h}(t, S, 0, \lambda) = 0$. Then, for $\gamma > 0$, we find for the optimal trading rate the explicit representation

$$\nu^*(t, S, q, \lambda) = \left(\frac{-\phi [K - S]^+ + \eta \partial_\lambda \tilde{h}(t, S, q, \lambda) - \partial_q \tilde{h}(t, S, q, \lambda)}{(\gamma + 1) \lambda \kappa} \right)^{1/\gamma},$$

and hence $\tilde{h}$ is finally a viscosity solution of the explicit PDE

$$\partial_t \tilde{h} + \frac{1}{2} \sigma^2 \partial_{SS}^2 \tilde{h} + \frac{\left(-\phi [K - S]^+ + \eta \partial_\lambda \tilde{h} - \partial_q \tilde{h} \right)^{(1+\gamma)/\gamma}}{((\gamma + 1) \lambda \kappa)^{1/\gamma}} \left(\frac{\gamma}{\gamma + 1} \right) - \alpha \lambda \partial_\lambda \tilde{h} = 0,$$

This explicit HJB equation is much more easy to solve than the implicit equation (5.B.2) based on the expected shortfall penalty (5.B.1). Moreover, the optimal trading rate $(\nu_t^*)_t$ is not deterministic anymore in this setting since it depends on S .

CHAPTER 6

Deep learning for high-dimensional stochastic optimal control without explicit solution

Several of the optimal strategies discussed in Chapter 5 lead to a high-dimensional optimal control which does not admit an explicit representation. In this case, most of the recent literature on deep learning for solving high-dimensional partial differential equations do not apply. This Chapter 6 therefore introduces the GPI-PINN algorithm, a novel numerical scheme for solving continuous-time stochastic optimal control problems in high dimensions when the optimal control does not admit an explicit solution. Combining Physics-Informed Neural Networks with an Actor-Critic structure built upon the Generalized Policy Iteration technique, this successive deep learning algorithm employs two separate neural networks to approximate both the value function and the multidimensional optimal control. This way, the GPI-PINN algorithm manages to achieve a global approximation of the optimal solution across all time and space, which can be evaluated online rapidly. The optimality and convergence of the scheme are demonstrated theoretically and its accuracy and efficacy are shown empirically based on two numerical examples. In particular, we generalize the standard Almgren-Chriss model arising from optimal liquidation by allowing for a price impact model with fully nonlinear temporary and permanent impact functions and by considering a multidimensional setting with numerous co-integrated assets.

Dupret, J. L., & Hainaut, D. (2024). Deep learning for high-dimensional continuous-time stochastic optimal control without explicit solution. *To be submitted.*

6.1 Introduction

A lot of literature has emerged recently for solving high-dimensional partial differential equations based on deep learning, see *e.g.* Han et al. (2018), Sirignano and Spiliopoulos (2018), Han et al. (2017), Beck et al. (2021), Blechschmidt and Ernst (2021). These algorithms can therefore be used to solve continuous-time (CT) stochastic (Gaussian) optimal control problems whenever the optimal control admits an explicit solution in terms of the value function. Indeed, in such cases, the well-known Hamilton-Jacobi-Bellman (HJB) equation reduces to a parabolic second-order PDE. One of the most promising numerical schemes is based on Physics-Informed Neural Networks (PINNs), as introduced in Raissi et al. (2019) and which is quite different from standard supervised machine learning techniques since instead of relying purely on data, it uses the physical properties of the PDE itself to guide the training process. PINNs indeed generate approximate solutions to PDEs by training a neural network to minimize a loss function consisting of terms representing the misfit of the initial and boundary conditions along points sampled from the boundary of the space-time domain, as well as the PDE residual at selected points in the interior of the domain. We refer in particular to Sirignano and Spiliopoulos (2018) and Blechschmidt and Ernst (2021), where PINNs are applied for solving such CT explicit stochastic optimal control problems in high dimensions. However, when the optimal control does not admit an explicit representation, one then has to simultaneously solve an optimization problem with respect to the control as well as a (second-order) PDE with respect to the value function, as we will see in Section 6.2.2. These implicit optimal control problems cannot be solved anymore using directly the deep learning techniques mentioned above and therefore require a specifically-designed successive approximation algorithm. For low-dimensional implicit optimal controls, a classical solution from the reinforcement learning literature is to use the (Generalized) Policy Iteration algorithm, which is a successive approximation algorithm for both the value function and the optimal control, see Jacka and Mijatović (2017), Sutton and Barto (2018) and Algorithm 4. In particular, we highlight the Actor-Critic methods from Werbos (1992), which have a separate memory structure to explicitly represent the optimal control independently of the value function and which rely on this (G)PI algorithm. However, these numerical schemes based on (G)PI usually difficulty scale with the problem's dimension as the optimization and the PDE both have to be solved for every point in a finite grid. Our algorithm should hence be meshfree, which is key since meshes become infeasible in higher dimensions.

Therefore, the numerical scheme proposed in this chapter aims at solving CT stochastic optimal control in high dimensions, even when the control cannot be expressed explicitly, and is inspired by both the deep learning and the reinforcement learning theory. More precisely, the so-called GPI-PINN algorithm is obtained by

combining PINNs with an Actor-Critic structure built upon the GPI algorithm. This successive GPI-PINN scheme is able to approximate both the value function and the high-dimensional optimal control using two separate neural networks (NNs), that we will train based on simulated data from the space-time domain. This provides a global approximation of the control and value function available across all time and space, which can then be evaluated online very fast. By imposing an early termination condition in the training of the GPI-PINN algorithm, we can prove the optimality and convergence guarantees of our numerical scheme towards the true solution of the value function and optimal control.

Finally, based on two detailed numerical examples, we show empirically the generality, accuracy and efficacy of the proposed GPI-PINN algorithm for solving implicit stochastic optimal control problems in high dimensions. First, we compare and confirm the performance and accuracy of our GPI-PINN algorithm on a classical linear-quadratic Gaussian control example for which the solution is known up to Monte-Carlo simulations. The second example arises from optimal liquidation in finance where an agent has to sell a large amount of shares and yet minimize price movements generated by her own trades, see Chapter 5 and the review of Carlea et al. (2015). In a first step, we generalize the standard Almgren and Chriss (2001) and Almgren (2003) models by assuming nonlinear (power-law) temporary and permanent impact functions as it prevents from having an explicit solution for the optimal control. We then extend this fully nonlinear optimal liquidation setting in higher dimensions by considering a large number of co-integrated assets, which interact and influence each other. This indeed leads to a high-dimensional CT implicit stochastic optimal control problem that can be handled efficiently and accurately using our GPI-PINN Algorithm.

The contributions of this article are therefore threefold. We first introduce a new successive deep-learning approximation algorithm able to solve high-dimensional CT stochastic optimal control problems when the control does not admit an explicit solution. We then prove theoretically the uniform convergence of this so-called GPI-PINN algorithm toward the true optimal solutions. Finally, we show empirically the efficacy and accuracy of the proposed numerical scheme on several examples. In particular, we generalize the standard Almgren-Chriss model arising from optimal liquidation problems by allowing for a fully nonlinear temporary and permanent impact functions and by considering a multidimensional setting with numerous co-integrated assets, as it leads to an implicit optimal control problem that can be effectively handled by our GPI-PINN algorithm.

The organization of the chapter is as follows. In Section 6.2, we review the literature and the theoretical foundations of implicit and explicit stochastic optimal controls. In Section 6.3, we introduce the GPI-PINN algorithm with early termination and prove its optimality and convergence. Section 6.4 lays out the neural

network implementation used in the rest of the chapter. Section 6.5 performs a comparative example for which a solution is known up to Monte-Carlo simulations and Section 6.6 shows how the GPI-PINN algorithm can be used to solve complex optimal liquidation strategies. Section 6.7 concludes.

6.2 Stochastic optimal control

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a complete filtered probability space on which is defined a d -dimensional Brownian motion $W = (W_t)_{t \geq 0}$. We fix $T < \infty$ and define $Q = [0, T] \times \mathbb{X}$ with state space $\mathbb{X} \subseteq \mathbb{R}^n$ and $\bar{Q} = [0, T] \times \mathbb{X}$. We consider a fixed final-time CT stochastic (Gaussian) optimal control problem of the form

$$\begin{cases} \underset{u \in \mathcal{A}}{\text{minimize}} & V^u(t, x) = \mathbb{E} \left[\phi(T, X_T^u) + \int_t^T \ell(\tau, X_\tau^u, u_\tau) d\tau \mid X_t^u = x \right], \\ \text{subject to} & dX_\tau^u = f(\tau, X_\tau^u, u_\tau) d\tau + \sigma(\tau, X_\tau^u, u_\tau) dW_\tau, \end{cases} \quad (6.2.1)$$

for all $(t, x) \in Q$. Here $X^u : [0, T] \times \Omega \rightarrow \mathbb{X}$ is the state process, $u : Q \times \Omega \rightarrow \mathbb{U} \subseteq \mathbb{R}^m$ is the control process and $\mathbb{U}$ is a closed subset of $\mathbb{R}^m$. $\mathcal{A}$ denotes the set of admissible controls, which consists of all progressively measurable m -dimensional processes $u = (u_t)_{t \in [0, T]}$ with finite $\mathbb{L}^2(\Omega \times [0, T])$ -norm. The drift and diffusion coefficients $f : Q \times \mathbb{U} \rightarrow \mathbb{R}^n$ and $\sigma : Q \times \mathbb{U} \rightarrow \mathbb{R}^{n \times d}$ are assumed to be continuous and satisfy the Lipschitz conditions

$$\begin{aligned} |f(t, x, u) - f(t, y, u)| + |\sigma(t, x, u) - \sigma(t, y, u)| &\leq K|x - y|, \\ |f(t, x, u)| + |\sigma(t, x, u)| &\leq K(1 + |x| + |u|), \end{aligned} \quad (6.2.2)$$

for K independent of (t, x, y, u) . From Touzi (2012), for each admissible control $u \in \mathcal{A}$, the controlled SDE (6.2.1) has then a unique $\mathbb{F}$ -adapted solution X^u , for all $X_0 \in \mathbb{L}^2(\mathcal{F}_0, \mathbb{P})$ such that for some constant C :

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s^u|^2 \right] < C (1 + \mathbb{E}[|X_0|^2]) e^{Ct} \quad \text{for all } t \in [0, T].$$

Finally, $\phi : \mathbb{X} \rightarrow \mathbb{R}$ is the terminal cost at T and $\ell : Q \times \mathbb{U} \rightarrow \mathbb{R}$ is a continuous running cost. They satisfy the following quadratic growth condition for some constant K independent of (t, x, u) , $|\ell(t, x, u)| + |\phi(x)| \leq K(1 + |u|^2 + |x|^2)$. This quadratic growth on ℓ and ϕ ensures that $V^u(t, x)$ is well-defined for all $u \in \mathcal{A}$. We consider in this work Markovian controls, *i.e.* processes $u \in \mathcal{A}$ that can be written in the form $u_t = u(t, X_t^u)$ for each $t \in [0, T]$ and for some measurable map u from Q into $\mathbb{U}$. We also denote $u_t^* = u^*(t, X_t^{u^*})$ the optimal closed-loop feedback control minimizing the cost V^u (6.2.1), with $u^* : Q \rightarrow \mathbb{U}$ measurable. Standard notation in dynamic programming defines the value function $V : \bar{Q} \rightarrow \mathbb{R}$ as the optimal

cost-to-go (6.2.1) starting at $(t, x) \in Q$, i.e.

$$V(t, x) := \inf_{u \in \mathcal{A}} V^u(t, x) = \mathbb{E} \left[\phi(T, X_T^{u^*}) + \int_t^T \ell(\tau, X_\tau^{u^*}, u^*(\tau, X_\tau^{u^*})) d\tau \mid X_t^{u^*} = x \right], \quad (6.2.3)$$

with terminal condition $V(T, x) = \phi(T, x)$. Let $\mathcal{S}_n$ the set of all $n \times n$ symmetric matrices with real coefficients. We next define the Hamiltonian by the map $H : Q \times \mathbb{R}^n \times \mathcal{S}_n \rightarrow \mathbb{R}$,

$$\mathcal{H}(t, x, p, \gamma, u) := \ell(t, x, u) + p^\top f(t, x, u) + \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x, u) \gamma]. \quad (6.2.4)$$

The minimized Hamiltonian is then defined as

$$H(t, x, p, \gamma) := \inf_{u \in \mathbb{U}} \mathcal{H}(t, x, p, \gamma, u) = \mathcal{H}(t, x, p, \gamma, u^*(t, x)), \quad (6.2.5)$$

and we have the following crucial theorem related to the value function V , see the review of Touzi (2012).

Theorem 6.1. *Assume the value function $V \in C^{1,2}(Q)$, let the minimized Hamiltonian H in (6.2.5) be continuous in (t, x) with $H(\cdot, \cdot, V_x, V_{xx}) < \infty$ and let $\ell(\cdot, \cdot, u)$ be continuous in (t, x) for all fixed $u \in \mathbb{U}$, then the value function V solves the following HJB PDE for every $(t, x) \in Q$,*

$$\begin{cases} V_t(t, x) + \inf_{u \in \mathbb{U}} \left\{ \ell(t, x, u) + [V_x(t, x)]^\top f(t, x, u) + \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x, u) V_{xx}(t, x)] \right\} = 0 \\ V(T, x) = \phi(x), \end{cases} \quad (6.2.6)$$

$$\Leftrightarrow \begin{cases} V_t(t, x) + \ell(t, x, u^*(t, x)) + [V_x(t, x)]^\top f(t, x, u^*(t, x)) + \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x, u^*(t, x)) V_{xx}(t, x)] = 0, \\ V(T, x) = \phi(x), \end{cases} \quad (6.2.7)$$

where we denote $V_t := \partial V / \partial t$, $V_x := (\partial V / \partial x_1, \dots, \partial V / \partial x_n)^\top$ the gradient of V and V_{xx} the Hessian of V .

If V satisfies Theorem 6.1, then it is said to be a classical solution of the HJB equation (6.2.6). We directly see that we can rewrite this HJB equation (6.2.6) as

$$\begin{cases} V_t(t, x) + H(t, x, V_x(t, x), V_{xx}(t, x)) = 0, \\ V(T, x) = \phi(x), \end{cases} \quad (6.2.8)$$

and we therefore have that the optimal control u^* must solve

$$u^*(t, x) := u^*(t, x; V_x, V_{xx}) = \arg \min_{u \in \mathbb{U}} \mathcal{H}(t, x, V_x(t, x), V_{xx}(t, x), u). \quad (6.2.9)$$

We now give a crucial verification theorem, the proof of which is outlined in Touzi (2012).

Theorem 6.2. *Let $\hat{V} \in C^{1,2}(Q) \cap C(\bar{Q})$ be a solution of (6.2.7). If there exists $\hat{u}(t, x)$ for each $(t, x) \in Q$ such that¹*

$$\hat{u}(t, x) \in \arg \min_{u \in \mathbb{U}} \mathcal{H}(t, x, \hat{V}_x(t, x), \hat{V}_{xx}(t, x), u), \quad (6.2.10)$$

then $\hat{V}(t, x) = V(t, x)$ for all $(t, x) \in \bar{Q}$ and $\hat{u}(t, x)$ must be an optimal (Markov) control $u^*(t, x)$ for all $(t, x) \in Q$.

From this verification theorem, we will propose in this chapter as candidate $\hat{V}$ a NN-parametrized value function $V(t, x; \theta)$ on $\bar{Q}$ chosen to be $C^{1,2}(Q) \cap C(\bar{Q})$ and whose neural weights θ are estimated in such a way that the PDE (6.2.7) is satisfied. We will then aim to find a $\hat{u}$ minimizing (6.2.10). This control $\hat{u}$ will also assumed to be parameterized by a (second) NN $u(t, x; \omega)$ on Q and the weights ω chosen so as to minimize (6.2.10). Note that under the assumption that the coefficients $f(\cdot)$, $\sigma(\cdot)$ and $\ell(\cdot)$ are C^1 in u , equation (6.2.9) implies

$$\mathcal{H}_u(t, x, V_x(t, x), V_{xx}(t, x), u^*(t, x)) := \left. \frac{\partial}{\partial u} \mathcal{H}(t, x, V_x(t, x), V_{xx}(t, x), u) \right|_{u=u^*(t, x)} = 0. \quad (6.2.11)$$

Moreover, from Theorem 6.2, we see that finding a $\hat{u}$ such that

$$\left. \frac{\partial}{\partial u} \mathcal{H}(t, x, \hat{V}_x(t, x), \hat{V}_{xx}(t, x), u) \right|_{u=\hat{u}(t, x)} = 0, \quad (6.2.12)$$

provides a good candidate for the optimal control, even if this condition is not sufficient for ensuring the optimality of u^* . Since $\hat{V}$ should be $C^{1,2}(Q) \cap C(\bar{Q})$ in the verification Theorem 6.2, choosing $\hat{V}(t, x) = V(t, x; \theta)$ where $V(\cdot; \theta)$ is a NN belonging to $C^{1,2}(Q) \cap C(\bar{Q})$ should provide a natural candidate for the value function. Choosing two-times differentiable activation functions in our NN ensures that $V(\cdot; \theta)$ is $C^2(\bar{Q})$ as the intersection of two-times differentiable activation functions keeps the output function $V(\cdot; \theta)$ in $C^2(\bar{Q})$. Therefore, choosing for example the tanh, sigmoid, softplus or linear activation functions implies that the approximated NN value function $V(\cdot; \theta)$ is smooth and can therefore solve for a classical solution of the PDE (6.2.6). Moreover, the choice of a NN approximation for the value function is also justified by the universal approximation theorem (see Lemma 6.1 below) stating that any continuous (and hence any $C^2(\bar{Q})$) value function on a compact domain can be approximated by a suitable NN. The following notation will also be useful in Section 6.3 when proving the convergence and optimality of the proposed algorithm :

$$\mathcal{H}'(t, x, V_x(t, x), V_{xx}(t, x), u(t, x)) := \mathcal{H}(t, x, V_x(t, x), V_{xx}(t, x), u(t, x)) + V_t(t, x). \quad (6.2.13)$$

¹or similarly such that $\hat{V}_t(t, x) + \mathcal{H}(t, x, \hat{V}_x(t, x), \hat{V}_{xx}(t, x), \hat{u}(t, x)) = 0$, see (6.2.7).

6.2.1 Explicit optimal control

In CT stochastic optimal control problems, the minimization problem (6.2.9) in u can often be simplified so as to express the optimal control u^* explicitly as a function of t, x, V_x and V_{xx} . Therefore, in this case, the minimized Hamiltonian $H(t, x, V_x, V_{xx})$ can also be obtained explicitly as a function of (t, x, V_x, V_{xx}) , as shown *e.g.* in Example 1 below with equation (6.2.16). The PDE (6.2.8) is then directly solved for V since it does not depend anymore on the optimal control. This value function $V(t, x)$ obtained from the HJB equation is next plugged back into the formula for $u^*(t, x)$ so as to also derive the optimal control explicitly. This two-step method (with first the derivation of V and then of u^*) is illustrated in the following example.

Example 1 : We consider a simple case with an affine stochastic control and convex running cost in $u \in \mathbb{U} = \mathbb{R}^m$. The system dynamic is given by

$$dX_t^u = f(t, X_t^u)u_t dt + \sigma(t, X_t^u)dW_t, \quad X_0 = x_0.$$

where $f : Q \rightarrow \mathbb{R}^{n \times m}$, $\sigma : Q \rightarrow \mathbb{R}^{n \times d}$. The running cost is assumed to be of the form $\ell(t, x, u) = h(t, x) + u^\top \mathbf{R}u$, for some convex function $h : Q \rightarrow \mathbb{R}$ and some positive definite weight matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$. Then, the Hamiltonian (6.2.4) becomes

$$\mathcal{H}(t, x, p, \gamma, u) = h(t, x) + u^\top \mathbf{R}u + p^\top f(t, x)u + \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x) \gamma]. \quad (6.2.14)$$

The minimization problem (6.2.9) for u^* with $p = V_x$ amounts to compute the derivative of $\mathcal{H}$ with respect to u and set it to zero

$$\mathbf{0}_m = \mathcal{H}_u(t, x, V_x(t, x), V_{xx}(t, x), u^*(t, x)) = 2\mathbf{R}u^*(t, x) + f^\top(t, x)V_x(t, x).$$

Solving for u^* yields the optimal feedback control law in explicit form

$$u^*(t, x) = -\frac{1}{2}\mathbf{R}^{-1}f^\top(t, x)V_x(t, x), \quad (6.2.15)$$

and substituting u^* back in (6.2.14), we also obtain the minimized Hamiltonian explicitly

$$\begin{aligned} H(t, x, V_x(t, x), V_{xx}(t, x)) &= h(t, x) - \frac{1}{4}V_x^\top(t, x)f(t, x)\mathbf{R}^{-1}f^\top(t, x)V_x(t, x) \\ &\quad + \frac{1}{2}\text{Tr} [\sigma \sigma^\top(t, x)V_{xx}(t, x)]. \end{aligned} \quad (6.2.16)$$

We then obtain for all $(t, x) \in Q$ the following HJB PDE satisfied by the value function V :

$$V_t(t, x) + h(t, x) - \frac{1}{4}V_x^\top(t, x)f(t, x)\mathbf{R}^{-1}f^\top(t, x)V_x(t, x) + \frac{1}{2}\text{Tr} [\sigma \sigma^\top(t, x)V_{xx}(t, x)] = 0, \quad (6.2.17)$$

with terminal condition $V(T, x) = \phi(x)$. We hence see that this HJB PDE (6.2.17) can be solved independently of the control u . Once the value function V is solved, we can plug its space derivative into (6.2.15) and obtain the optimal control u^* . Note that if the diffusion coefficient $\sigma(\cdot)$ also depends (linearly) on the control u , then the optimal control will be a function of the diagonal elements of the Hessian V_{xx} in the denominator.

6.2.1.1 Literature review

Solving the parabolic PDE (6.2.17) for the value function arising from CT stochastic optimal control often requires numerical methods, especially in higher dimensions. Recently, many numerical schemes based on deep learning have emerged and have been proposed for solving such kind of PDEs. Indeed, as deep learning techniques for solving PDEs limit the increased complexity when considering high-dimensional problems (see *e.g.* Hutzenthaler et al. (2020)), they appear the best suited for solving such explicit CT stochastic optimal control in high dimensions. Moreover, they offer the possibility to build complex functions $f(t, x; \theta)$ that can directly approximate the value function and/or the control across all time and space. Once the set of parameters θ is estimated offline, these functions can then be rapidly evaluated online for every (t, x) of the domain. Such deep learning techniques can be divided into two main approaches, as investigated in the review paper of Blechschmidt and Ernst (2021). First, we have the Physics-Informed Neural Networks (PINNs) introduced in Raissi et al. (2019), which are quite different from standard supervised machine learning techniques since instead of relying purely on data, it uses the physical properties of the PDE itself to guide the training process. They indeed generate approximate solutions to PDEs by training a neural network to minimize a loss function consisting of terms representing the misfit of the initial and boundary conditions along points sampled from the boundary of the space-time domain, as well as the PDE residual at selected points in the interior of the domain, see equation (6.3.1) below. The strength of PINNs lies in their flexibility in that they can be applied to a great variety of challenging PDEs, see *e.g.* Sirignano and Spiliopoulos (2018) with the DGM algorithm, Berg and Nyström (2018) or Glau and Wunderlich (2022), see also Bai et al. (2022) or Hainaut (2023) in finance and Mowlavi and Nabi (2023) in the context of optimal control. Moreover, PINNs provide a global solution of the PDE across all time and space (t, x) in the domain, which is much more suitable for real-time applications. The second set of methods consists in algorithms based on the Feymann-Kac formula for solving (Kolmogorov) linear parabolic second-order PDEs, as introduced in Beck et al. (2018) and Berner et al. (2020). Such Feymann-Kac approach connects the solution of the Komogorov PDE to the expectation of a stochastic process, which appears to work extremely well in high dimensions. This approach can also be extended to semi-linear (or even fully nonlinear) PDEs in high dimensions through a nonlinear Feymann-Kac representation, which allows to make the correspondence between nonlinear PDEs and

backward stochastic differential equations (BSDEs), see the deep BSDE solver of Han et al. (2017), Han et al. (2018) and its application for solving high-dimensional HJB PDEs. Many extensions of this BSDE solver are proposed in the literature, see *e.g.* Beck et al. (2021), Beck et al. (2019), Chan-Wai-Nam et al. (2019). Note that this second set of methods provides local solutions and hence focus on computing the value of the PDE solution at a single point in time, on the contrary to the global solution provided by PINNs. Of course, some techniques in the literature do not belong to any of these two categories such as the methods of Nakamura-Zimmerer et al. (2021) or Ji et al. (2022) using the Pontryagin's maximum principle which are semi-local and valid for a single initial state. Therefore, deviations of the system from the optimal trajectory require re-computation of the solution.

6.2.2 Implicit optimal control

Unlike the previous section on explicit controls, we are especially interested in this Chapter 6 in the case where the optimal control u^* cannot be obtained explicitly as a function of (t, x, V_x, V_{xx}) . Instead, it is only available implicitly as soon as the minimization (6.2.9) does not admit an analytical solution (*e.g.* non-convex Hamiltonian $\mathcal{H}$ in u). We then have to solve simultaneously the HJB equation (6.2.6) for the value function V and the minimization problem (6.2.9) for the optimal control u^* . Numerical methods exist based on recursive schemes for approximating u^* (see Algorithm 4 below), but these schemes are already computationally intensive even for low-dimensional problems. Hence, the curse of dimensionality² that typically appears for high-dimensional PDEs is therefore exacerbated by the extra simultaneous minimization step required when no explicit expression is available for the optimal control u^* . This raises the need of alternative numerical techniques based on deep learning for tackling implicit CT stochastic optimal control problems, especially in high dimensions, as shown in the following example.

Example 2 : We now consider the following stochastic dynamic for the state

$$dX_t^u = f(t, X_t^u) \sin(u_t) dt + \sigma(t, X_t^u) dW_t, \quad X_0 = x_0,$$

for some function $f : Q \rightarrow \mathbb{R}^{n \times m}$, $\sigma : Q \rightarrow \mathbb{R}^{n \times d}$ and where the running cost $\ell(t, x, u) = h(t, x) + u^\top \mathbf{R}u$, as in Example 1. The Hamiltonian then becomes

$$\mathcal{H}(t, x, p, \gamma, u) = h(t, x) + u^\top \mathbf{R}u + p^\top f(t, x) \sin(u) + \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x) \gamma]. \quad (6.2.18)$$

The minimization problem (6.2.9) for u^* with $p = V_x$ and $\gamma = V_{xx}$ implies

$$\mathbf{0}_m = \mathcal{H}_u(t, x, V_x(t, x), V_{xx}(t, x), u^*(t, x)) = 2\mathbf{R}u^*(t, x) + V_x^\top(t, x) f(t, x) \cos(u^*(t, x)), \quad (6.2.19)$$

²This term was in fact first formulated by Bellman in the 1950s in the setting of optimal control problems.

which does not admit an explicit solution for u^* . Therefore, we cannot proceed as in Example 1 and we must solve the optimal control problem numerically since equation (6.2.19) and the HJB PDE (6.2.7) should now be solved simultaneously.

6.2.2.1 Literature review

The literature on CT stochastic optimal control does not provide lots of insights for solving implicit problems of the form (6.2.1), especially when the dimension is high. Indeed, most of the available results based on deep learning are related to explicit problems, as seen in the previous section. Among the few solutions available, we can cite Han and E (2016) or Ji et al. (2022), which both propose local methods and hence do not provide a global solution of the implicit control problem for all points in the space-time domain. Moreover, they only parameterize the optimal control $u^*(t, x)$ by a NN and not the value function, which thus cannot be evaluated online rapidly. Fortunately, we can borrow from the literature on reinforcement learning (mainly developed for discrete-time problems) some insights on how to solve these implicit CT stochastic optimal control problems. One of the most classical ways is to use the Howard or Policy Iteration (PI) algorithm, originally derived in Howard (1960) for discrete-time reinforcement learning applications. This algorithm defines a successive approximation algorithm that can also be applied for CT stochastic optimal control problems, as highlighted in Algorithm 4 below.

As in equation (6.2.1), the approximated value function $V(t, x)$ for every $(t, x) \in \bar{Q}$ is denoted at step k of Algorithm 4 by $V^{u^{(k)}}(t, x)$ so as to highlight the dependence on the control $u^{(k)}(t, x)$, computed at the previous iteration. It can be shown from Peyrl et al. (2005) and Jacka and Mijatović (2017) that $V^{u^{(k+1)}}(t, x) < V^{u^{(k)}}(t, x)$ with $\lim_{k \rightarrow \infty} V^{u^{(k)}}(t, x) = V(t, x)$ for all $(t, x) \in \bar{Q}$ and that $\lim_{k \rightarrow \infty} u^{(k)}(t, x) = u^*(t, x)$ for all $(t, x) \in Q$. For applying this successive algorithm, we need to solve the PDE (6.2.20), using *e.g.* finite difference schemes. From this finite grid, the optimization problem (6.2.21) then needs to be solved for every grid point using approximation methods, which hence difficulty scale with the problem's dimension: BFGS and its extended techniques, gradient descent methods, Sequential least squares programming, and so on (see the review of Greif (2017)). Note that many more advanced methods like in Jensen and Smears (2013) or in Debrabant and Jakobsen (2013) have Howard's algorithm included. In the original PI Algorithm 4, we have to completely solve (6.2.20) before fully evaluating (6.2.21). This is however not necessary. Indeed, PI belongs to the Generalized Policy Iteration (GPI) techniques, which is a family of reinforcement learning techniques introduced in Sutton and Barto (2018) with PI at one extreme and value iteration at the other extreme. In GPI, at each step, one does not completely evaluate the value $V^{u^{(k)}}$ for a given control u^k , but only updates the current cost estimate towards that value. Likewise, one does not fully update the control policy to the greedy control

Algorithm 4 PI for CT stochastic optimal control

Choose an initial control law $u^{(0)}(\cdot) \in \mathcal{A}$.

Set any small positive number ε and let $k = 0$.

while $\max_{(t,x) \in \bar{Q}} |V^{u^{(k)}}(t,x) - V^{u^{(k-1)}}(t,x)| \geq \varepsilon$ **do**

1. Find $V^{u^{(k)}}(t,x)$ for all $(t,x) \in \bar{Q}$ by solving the following PDE for the already known control $u^{(k)}$:

$$\begin{aligned} V_t^{u^{(k)}}(t,x) + \ell(t,x, u^{(k)}(t,x)) + \left[V_x^{u^{(k)}}(t,x) \right]^\top f(t,x, u^{(k)}(t,x)) \\ + \frac{1}{2} \text{Tr} \left[\sigma \sigma^\top(t,x, u^{(k)}(t,x)) V_{xx}^{u^{(k)}}(t,x) \right] = 0 \end{aligned} \quad (6.2.20)$$

with terminal condition $V^{u^{(k)}}(T,x) = \phi(x)$.

2. Compute the control $u^{(k+1)}(t,x)$ at step $k+1$ from the known functional $V^{u^{(k)}}(t,x)$ by solving for every $(t,x) \in Q$:

$$u^{(k+1)}(t,x) = \arg \min_{u \in \mathbb{U}} \left\{ \ell(t,x,u) + \left[V_x^{u^{(k)}}(t,x) \right]^\top f(t,x,u) + \frac{1}{2} \text{Tr} \left[\sigma \sigma^\top(t,x,u) V_{xx}^{u^{(k)}}(t,x) \right] \right\} \quad (6.2.21)$$

$k = k + 1$

end while

Output $V^{u^{(k)}}(t,x)$ for every $(t,x) \in \bar{Q}$ and $u^{(k)}(t,x)$ for every $(t,x) \in Q$.

defined as $u^{(k+1)}(t,x) := \arg \min_{u \in \mathbb{U}} \mathcal{H}(t,x, V_x^{u^{(k)}}(t,x), V_{xx}^{u^{(k)}}(t,x), u)$ for the new value estimate at step k , but only updates the control towards this greedy control $u^{(k+1)}(t,x)$. In value iteration, we only perform one iteration at each step while in PI we wait until full convergence is reached at each step in (6.2.20)-(6.2.21). Numerous other reinforcement algorithms (SARSA, Q-learning, TD(λ), etc.) are presented in Sutton and Barto (2018) or in Bertsekas (2023) and can be used for solving discrete-time (deterministic and stochastic) optimal control problems. Among these algorithms, we highlight the Actor-Critic methods (or adaptive dynamic programming from Werbos (1992)), which have a separate memory structure to explicitly represent the optimal control u^* independently of the value function V . Most of Actor-Critic methods then rely on a (G)PI scheme as in Algorithm 4 and were initially developed for stationary discrete-time stochastic problems (with infinite horizon). As we will show in Section 6.3, our methodology extends these methods in CT for stochastic control problems with finite-time horizon.

Moreover, the above reinforcement learning methods are not suitable as is for

solving high-dimensional CT stochastic control problems of the form (6.2.1) when the optimal control is only available implicitly. On the one hand, these algorithms do not solve the fact mentioned above that implicit control problems difficulty scale when increasing the number of dimensions n (or m) and are therefore especially prone to suffer from the curse of dimensionality. This can be overcome in discrete-time by combining these algorithms with function approximations based on deep learning (and NNs), see the review of Li (2017). Indeed, they either aim at estimating only the value function (or action-value function), as in Chapter 3 of Bertsekas (2023), or at estimating both this value function and the optimal control in a separate way using an enhanced Actor-Critic method with one NN for the control and one NN for the value function at each time step (but no global solution), see Huré et al. (2021) and Bachouch et al. (2022). Proofs of convergence in the stochastic discrete-time case with finite horizon are given in Huré et al. (2021). On the other hand, for the CT case of interest in this chapter, only deterministic dynamics with infinite horizon are considered in the literature. Indeed, with the seminal paper of Abu-Khalaf and Lewis (2005), a second stream of literature tackles CT optimal control problems with deterministic nonlinear dynamic and infinite horizon using an Actor-Critic structure in an offline way. Moreover, in Vrabie and Lewis (2008) and Vrabie et al. (2009) is developed an online PI algorithm for deterministic CT systems which converges to the optimal control solution without making explicit use of any knowledge on the internal dynamics of the system. The algorithm is based on sequential updates of the critic (policy evaluation) and actor (policy improvement) neural networks: while one NN is tuned the other one remains constant. We will see in Section 6.3 how this Actor-Critic structure based on separate NNs for the value function and the optimal control can also help to solve CT optimal control problems with stochastic dynamic.

Therefore, we now aim at developing a deep learning algorithm able to approximate both the optimal control and the value function in our CT stochastic optimal control problem (6.2.1) with finite horizon, which will be specifically designed to perform well in a high-dimensional setting with implicit control. Our method, inspired by both the reinforcement learning theory and the PINN literature, will provide a global approximation of the (m -dimensional) optimal control and of the value function available across all time and space, which can then be evaluated online very fast. This will be achieved more precisely by combining PINNs from Section 6.2.1 along with an Actor-Critic structure based on the (G)PI scheme of Algorithm 4. Our proposed algorithm will also perform well for explicit controls in high dimensions as it avoids the unstable computation of the ratio V_x/V_{xx} when the diffusion coefficient $\sigma(\cdot)$ is a function of the control u .

6.3 Deep learning for implicit CT stochastic optimal control

6.3.1 Standard PINN

As already mentioned in Section 6.2.1.1, PINNs are able to approximate the value function (6.2.3) when the optimal control u^* is given explicitly as a function of (t, x, V_x, V_{xx}) . PINNs are indeed built such as to minimize the residual of the HJB equation (6.2.8) in its interior domain along with the residual of its initial and/or boundary conditions. More precisely, the loss function is defined

$$\mathcal{L}(\theta) = \left\| \partial_t V(t, x; \theta) + H(t, x, V_x(t, x; \theta), V_{xx}(t, x; \theta)) \right\|_{Q, \mu_1}^2 + \left\| V(T, x; \theta) - \phi(x) \right\|_{\mathbb{X}, \mu_2}^2, \quad (6.3.1)$$

where $\|g(y)\|_{\mathcal{Y}, \mu}^2 = \int_{\mathcal{Y}} |g(y)|^2 \mu(y) dy$ and $\mu(y)$ a positive probability density on $\mathcal{Y}$, and where $V(t, x; \theta)$ is a NN parameterized by θ and approximating the value function (6.2.3). As explained in the DGM algorithm of Sirignano and Spiliopoulos (2018), the loss function $\mathcal{L}(\theta)$ measures how well the function $V(t, x; \theta)$ satisfies the HJB equation (6.2.8). If $\mathcal{L}(\theta) = 0$, then $V(t, x; \theta)$ is solution of the HJB equation. Therefore, the goal is to find a set of parameters θ such that the function $V(t, x; \theta)$ minimizes the error $\mathcal{L}(\theta)$. Hence, a θ which minimizes $\mathcal{L}(\theta)$ produces a reduced-form model $V(t, x; \theta)$ which approximates $V(t, x)$ for all $(t, x) \in \bar{Q}$, see Sirignano and Spiliopoulos (2018), Jiang et al. (2023) and De Ryck and Mishra (2022) for convergence results. Estimating θ by minimizing directly $\mathcal{L}(\theta)$ is infeasible when the dimension n is large since the integral over $\mathbb{X}$ is intractable computationally. However, using stochastic gradient descent (SGD) (or mini-batch gradient descent) with machine learning approach limits the curse of dimensionality by drawing a sequence of time and space points (according to $\mu_1(Q)$ and $\mu_2(\mathbb{X})$) while satisfying the dynamic (6.2.1). Note that potential boundary conditions on the HJB equation inferred from the physics of the problem can also be considered easily by adding a third residual term $\left\| V(t, x; \theta) - g(t, x) \right\|_{\partial Q, \mu_3}^2$ in (6.3.1), defined on the boundary $\partial Q := [0, T) \times \partial \mathbb{X}$. Such boundary conditions will be used in Section 6.6 on optimal liquidation problems.

In the explicit case of Section 6.2.1, once $V(t, x)$ is approximated by $V(t, x; \theta)$, the optimal control can then be derived directly from the value function, see equation (6.2.15). However, as shown above, this method cannot be applied when the control does not admit an explicit representation in terms of $V(t, x)$ since both the optimal control $u^*(t, x)$ and the value function $V(t, x)$ need to be approximated simultaneously. In this case, we know from (6.2.5) that the optimal control $u^*(t, x)$ should minimize the Hamiltonian $\mathcal{H}(t, x, V_x(t, x), V_{xx}(t, x), u)$ for $u \in \mathbb{U}$ while the

value function should satisfy the HJB equation (6.2.8). We can then try to combine the HJB equation (6.2.8) and the minimization (6.2.5) in a PINN setup so as to define the following loss function

$$\begin{aligned} \mathcal{L}(\theta, \omega) = & \left\| \partial_t V(t, x; \theta) + \mathcal{H}(t, x, V_x(t, x; \theta), V_{xx}(t, x; \theta), u(t, x; \omega)) \right\|_{Q, \mu_1}^2 \\ & + \left\| V(T, x; \theta) - \phi(x) \right\|_{\mathbb{X}, \mu_2}^2 + \int_Q \mathcal{H}(t, x, V_x(t, x; \theta), V_{xx}(t, x; \theta), u(t, x; \omega)) d\mu_1(t, x) \end{aligned} \quad (6.3.2)$$

where $V(t, x; \theta)$ again approximates the value function via a NN parametrized by θ and $u(t, x; \omega)$ is a second NN parameterized by ω which approximates the m -dimensional optimal control. However, this cannot work since the minimization of $\mathcal{H}(t, x, V_x, V_{xx}, u)$ in the last term of (6.3.2) should be performed with respect to the control u via ω (and not the value function V via θ), while the HJB equation in the two first terms should be solved with respect to V via θ (and not via ω). Clearly, this cannot be ensured by the loss function (6.3.2). To avoid that, we therefore need to sequentially solve the HJB equation with respect to the weights θ of the first NN $V(t, x; \theta)$ while keeping the weights ω of the optimal control NN constant, and then minimize the Hamiltonian with respect to the weights ω of the second NN $u(t, x; \omega)$ while keeping the weights θ of the value NN constant. The (G)PI Algorithm 4 with an Actor-Critic structure built upon the two NNs $V(\cdot; \theta)$ and $u(\cdot; \omega)$ will therefore provide the perfect framework for this successive optimization scheme.

6.3.2 Sequential PINN

We therefore first propose in Section 6.3.2.1 the PI-PINN Algorithm 5, which combines PINNs as developed above with an Actor-Critic structure based on the PI scheme of Algorithm 4. We will see that it leads to a global approximation across all time and space of both the m -dimensional optimal control and the value function, even in high dimensions and when there is no explicit solution of the optimal control. This PI-PINN Algorithm 5 is then enhanced in Section 6.3.2.2 with the GPI-PINN Algorithm 6, which implements a GPI scheme instead with early termination allowing for faster numerical convergence. We also derive optimality and theoretical convergence guarantees for this improved algorithm with Theorem 6.3.

6.3.2.1 PI-PINN Algorithm

We can directly see in Algorithm 5 below that we first minimize the loss function (6.3.4) with respect to θ so as to enforce the HJB equation (6.2.7) satisfied by the value function (even if at step k the control is not yet optimal). We then minimize the second loss function (6.3.6) with respect to ω so as to minimize the Hamiltonian (6.2.4) with respect to the control u in an average sense according

Algorithm 5 PI-PINN for implicit CT stochastic optimal control

Initialize arbitrary weights $\omega^{(0)}$ for the NN $u(\cdot; \omega^{(0)})$ and arbitrary weights $\theta^{(0)}$ for the NN $V(\cdot; \theta^{(0)})$. Set the tolerance to any small positive number ε and let $k = 0$.

while $\max_{(t,x) \in \bar{Q}} |V(t, x; \theta^{(k)}) - V(t, x; \theta^{(k-1)})| \geq \varepsilon$ **do**

1. For all $(t, x) \in \bar{Q}$, estimate the value NN $V(t, x; \theta^{(k+1)})$ for the known control $u(t, x; \omega^{(k)})$ at step k , by minimizing the following loss function $\mathcal{L}_1$ with respect to the weights θ :

$$\theta^{(k+1)} = \arg \min_{\theta} \mathcal{L}_1(\theta, \omega^{(k)}), \quad (6.3.3)$$

where

$$\begin{aligned} \mathcal{L}_1(\theta, \omega^{(k)}) &= \left\| \partial_t V(t, x; \theta) + \mathcal{H}(t, x, V_x(t, x; \theta), V_{xx}(t, x; \theta), u(t, x; \omega^{(k)})) \right\|_{Q, \mu_1}^2 + \left\| V(T, x; \theta) - \phi(x) \right\|_{\mathbb{X}, \mu_2}^2 \\ &= \mathbb{E}_{(t,x) \sim \mu_1(Q)} \left[\mathcal{H}' \left(t, x, V_x(t, x; \theta), V_{xx}(t, x; \theta), u(t, x; \omega^{(k)}) \right)^2 \right] + \mathbb{E}_{x \sim \mu_2(\mathbb{X})} [(V(T, x; \theta) - \phi(x))^2]. \end{aligned} \quad (6.3.4)$$

2. For all $(t, x) \in Q$, estimate the optimal control NN $u(t, x; \omega^{(k+1)})$ based on the value NN $V(t, x; \theta^{(k+1)})$ by solving the following minimization with respect to the weights ω :

$$\omega^{(k+1)} = \arg \min_{\omega} \mathcal{L}_2(\theta^{(k+1)}, \omega), \quad (6.3.5)$$

where

$$\begin{aligned} \mathcal{L}_2(\theta^{(k+1)}, \omega) &= \int_Q \mathcal{H}' \left(t, x, V_x(t, x; \theta^{(k+1)}), V_{xx}(t, x; \theta^{(k+1)}), u(t, x; \omega) \right) d\mu_1(t, x) \\ &= \mathbb{E}_{(t,x) \sim \mu_1(Q)} \left[\mathcal{H}' \left(t, x, V_x(t, x; \theta^{(k+1)}), V_{xx}(t, x; \theta^{(k+1)}), u(t, x; \omega) \right) \right] \end{aligned} \quad (6.3.6)$$

$k = k + 1$

end while

Output $V(t, x; \theta^{(k)})$ for every $(t, x) \in \bar{Q}$ and $u(t, x; \omega^{(k)})$ for every $(t, x) \in Q$.

to μ_1 . Note indeed from the notation (6.2.13) that finding the weights ω of the NN optimal control minimizing $\mathcal{H}(\cdot)$ or $\mathcal{H}'(\cdot)$ is equivalent and working with $\mathcal{H}'(\cdot)$ will help in the proof of Theorem 6.3. Alternatively, from condition (6.2.12), we can impose instead of the loss (6.3.6) that the derivative of the Hamiltonian (or equivalently $\mathcal{H}'$) with respect to u is equal to zero using the loss

$$\mathcal{L}_{2'}(\theta^{(k+1)}, \omega) = \left\| \mathcal{H}'_u \left(t, x, V_x(t, x; \theta^{(k+1)}), V_{xx}(t, x; \theta^{(k+1)}), u(t, x; \omega) \right) \right\|_{Q, \mu_1}^2. \quad (6.3.7)$$

where $\mathcal{H}'_u(t, x, V_x(t, x), V_{xx}(t, x), u(t, x; \omega)) := \frac{\partial}{\partial u} \mathcal{H}'(t, x, V_x(t, x), V_{xx}(t, x), u) \Big|_{u=u(t, x; \omega)}$. This does not ensure to find a global minimum but this may speed up convergence in some cases. We also see that we only need to train two NNs $V(t, x; \theta)$ and $u(t, x; \omega)$ with only two sets of parameters θ and ω so as to approximate the value function $V(t, x)$ for all $(t, x) \in \bar{Q}$ and the optimal control $u^*(t, x)$ for all $(t, x) \in Q$. This differs from Huré et al. (2021) and Bachouch et al. (2022) where two sets of parameters $(\theta^{(n)}, \omega^{(n)})$ need to be estimated at each discretized time t_n . Therefore, Algorithm 5 leads to a global approximation of the value function $V(t, x; \theta^{(k_{\max})})$ and the optimal control $u(t, x; \omega^{(k_{\max})})$ on the space-time domain, where $k_{\max}$ is the last epoch in Algorithm 5. Once the parameters θ and ω are estimated based on some generated dataset, we can then rapidly evaluate the optimal control and the value function online, making it suitable for real-time applications. Note that the loss functions (6.3.4) and (6.3.6) are estimated in practice by the following sample average procedure. Generate M_1 random points $\{(t_m^{(1)}, x_m^{(1)})\}_{m=1}^{M_1}$ from Q according to μ_1 and M_2 random points $\{(x_m^{(2)})\}_{m=1}^{M_2}$ from $\mathbb{X}$ according to μ_2 . Then, the loss functions (6.3.4) and (6.3.6) are estimated using

$$\begin{aligned} \widehat{\mathcal{L}}_1(\theta, \omega^{(k)}) &= \frac{1}{M_1} \sum_{m=1}^{M_1} \left(\mathcal{H}' \left(t_m^{(1)}, x_m^{(1)}, V_x(t_m^{(1)}, x_m^{(1)}; \theta), V_{xx}(t_m^{(1)}, x_m^{(1)}; \theta), u(t_m^{(1)}, x_m^{(1)}; \omega^{(k)}) \right) \right)^2 \\ &\quad + \frac{1}{M_2} \sum_{m=1}^{M_2} \left(V(t, x_m^{(2)}; \theta) - \phi(x_m^{(2)}) \right)^2, \end{aligned} \quad (6.3.8)$$

and

$$\widehat{\mathcal{L}}_2(\theta^{(k)}, \omega) = \frac{1}{M_1} \sum_{m=1}^{M_1} \mathcal{H}' \left(t_m^{(1)}, x_m^{(1)}, V_x(t_m^{(1)}, x_m^{(1)}; \theta^{(k)}), V_{xx}(t_m^{(1)}, x_m^{(1)}; \theta^{(k)}), u(t_m^{(1)}, x_m^{(1)}; \omega) \right). \quad (6.3.9)$$

Finally, in the loss function (6.3.4) (and in its empirical counterpart (6.3.8)), we can introduce a proportionality factor between the first residual term and the second boundary term. This hyper-parameter can be fine-tuned for improving the convergence of the algorithm.

6.3.2.2 GPI-PINN Algorithm

However, as already mentioned above in the PI Algorithm 4, it is not required to wait for full convergence in (6.3.3) and (6.3.5) to the exact solution. One can instead use a GPI scheme and only iterate a limited number of times towards the true solution, which leads to faster convergence rates in various settings, see Sutton and Barto (2018). We therefore suggest the following GPI-PINN Algorithm 6, which is built from Algorithm 5 but instead allows for early termination in the minimization of (6.3.3) and (6.3.5). Since parameter optimization is typically achieved in NN via gradient descent, we will use in practice SGD or mini-batches on the empirical losses (6.3.8) and (6.3.9), for successfully training the value NN $V(\cdot; \theta)$ and the optimal control NN $u(\cdot; \omega)$.

Therefore, in Algorithm 6 below, the gradient steps (6.3.11) and (6.3.12) are based in practice on SGD or mini-batches using the empirical formulae (6.3.8) and (6.3.9). Indeed, multiple theoretical analyses and experimental studies like Allen-Zhu et al. (2019), Du et al. (2019) or Jentzen et al. (2021) have shown that optimization algorithms like SGD are capable of finding the global minimum of the (non-convex) training objective for feedforward NNs in polynomial time, provided that this NN is over-parameterized. This justifies the following assumption.

Assumption 6.1. *Optimization algorithms like SGD or batch-SGD can attain the global minimum of the loss functions (6.3.4) and (6.3.6) for over-parameterized NNs in polynomial time.*

We then recall the universal approximation theorem for feedforward NNs of Hornik et al. (1989).

Lemma 6.1. *For any continuous function $F(y)$ defined on a compact set $\bar{Q}$, it is possible to construct a feedforward NN with one hidden layer that uniformly approximates $F(y)$ on $\bar{Q}$ with arbitrarily small error $\varepsilon > 0$.*

This lemma justifies the following simplifying assumption, which will drastically lighten the proof of convergence of Algorithm 6 in Theorem 6.3.

Assumption 6.2. *The NN approximation errors ε from Lemma 6.1 are ignored when proving the convergence of Algorithm 6 in Theorem 6.3.*

We now present the main theoretical result of the chapter showing that the value NN $V(t, x; \theta)$ and optimal control NN $u(t, x; \omega)$ both converge uniformly to an optimum. We consider a feed forward NN with one hidden layer of N neurons such that $V(t, x; \theta) = \sigma_{V_{out}} \left(\sum_{i=1}^N \beta_i^V \sigma_V \left(\alpha_{1,i}^V t + \sum_{j=1}^n \alpha_{j+1,i}^V x_j + c_j^V \right) \right) \in \mathbb{R}$, where σ_V is the value activation function, $\sigma_{V_{out}}$ the output activation function and $\theta \in \mathbb{R}^p$ the set of all parameters of the value NN with $p = N(n+2) + n$. For the optimal control,

Algorithm 6 GPI-PINN for implicit CT stochastic optimal control

Initialize arbitrary weights $\omega^{(0)}$ for the NN $u(\cdot; \omega^{(0)})$ and arbitrary weights $\theta^{(0)}$ for the NN $V(\cdot; \theta^{(0)})$. Set the tolerance to any small positive number ε and let $k = 0$. Choose suitable learning rates η_0, η_1, η_2 .

Phase 1 : Warm-up

while $\max_{(t,x) \in Q} \mathcal{H}'(t, x, \theta^{(k)}, \omega^{(k)}) \geq 0$ **do**

Update θ and ω using :

$$\{\theta^{(k+1)}, \omega^{(k+1)}\} = \{\theta^{(k)}, \omega^{(k)}\} - \eta_0 \frac{d\mathcal{L}_1(\theta^{(k)}, \omega^{(k)})}{d\{\theta^{(k)}, \omega^{(k)}\}} \quad (6.3.10)$$

$k = k + 1$

end while

Phase 2 : GPI-PINN with early termination

while $\max_{(t,x) \in \bar{Q}} |V(t, x; \theta^{(k)}) - V(t, x; \theta^{(k-1)})| \geq \varepsilon$ **do**

1. For every $(t, x) \in \bar{Q}$, estimate $V(t, x; \theta^{(k+1)})$ based on the previous control $u(t, x; \omega^{(k)})$ from step 2. Set $\theta^{(k+1)} = \theta^{(k)}$.

repeat

Update $\theta^{(k+1)}$ using

$$\theta^{(k+1)} = \theta^{(k+1)} - \eta_1 \frac{d\mathcal{L}_1(\theta^{(k+1)}, \omega^{(k)})}{d\theta^{(k+1)}} \quad (6.3.11)$$

until $\mathcal{H}'(t, x, \theta^{(k)}, \omega^{(k)}) \leq \mathcal{H}'(t, x, \theta^{(k+1)}, \omega^{(k)}) \leq 0$ for all $(t, x) \in Q$.

2. For every $(t, x) \in Q$, estimate $u(t, x; \omega^{(k+1)})$ based on the previous value function $V(t, x; \theta^{(k+1)})$ from step 1. Set $\omega^{(k+1)} = \omega^{(k)}$.

repeat

Update $\omega^{(k+1)}$ using

$$\omega^{(k+1)} = \omega^{(k+1)} - \eta_2 \frac{d\mathcal{L}_2(\theta^{(k+1)}, \omega^{(k+1)})}{d\omega^{(k+1)}} \quad (6.3.12)$$

until $\mathcal{H}'(t, x, \theta^{(k+1)}, \omega^{(k+1)}) \leq \mathcal{H}'(t, x, \theta^{(k+1)}, \omega^{(k)})$ for all $(t, x) \in Q$.

$k = k + 1$

end while

Output $V(t, x; \theta^{(k)})$ for every $(t, x) \in \bar{Q}$ and $u(t, x; \omega^{(k)})$ for every $(t, x) \in Q$.

we assume $u(t, x; \omega) = \sigma_{u_{out}} \left(\sum_{i=1}^N \beta_i^u \sigma_u \left(\alpha_{1,i}^u t + \sum_{j=1}^n \alpha_{j+1,i}^u x_j + c_j^u \right) \right) \in \mathbb{R}^m$ with σ_u the control activation function, $\sigma_{u_{out}}$ the output activation function and $\omega \in \mathbb{R}^p$ the set of all parameters of the control NN.

Theorem 6.3. *Let us assume $\ell(t, x, u)$ and $\phi(T, x)$ non-negative and that Assumptions 6.1 and 6.2 hold. For an arbitrary optimal control NN $u(t, x; \omega^{(0)})$ with weights $\omega^{(0)}$ such that $\sigma_{u_{out}}(\cdot)$ is locally bounded and for an arbitrary initial value NN $V(t, x; \theta^{(0)})$ with weight $\theta^{(0)}$ and with twice differentiable activation functions $\sigma_V(\cdot)$ and $\sigma_{V_{out}}(\cdot)$ such that $\sigma_{V_{out}}(\cdot) \geq 0$, if these two NNs are tuned with Algorithm 6, then $V(t, x; \theta^{(k)}) \rightarrow V(t, x)$ uniformly on $\bar{Q}$ and $u(t, x; \omega^{(k)}) \rightarrow u^*(t, x)$ uniformly on Q as $k \rightarrow \infty$.*

The proof is delegated in Appendix 6.A. However, when using the sampling procedure with loss functions (6.3.8) and (6.3.9), it can be intractable to check the values of $\mathcal{H}'(t, x, \omega, \theta)$ for every sampled points $\{(t_m^{(1)}, x_m^{(1)})\}_{m=1}^{M_1}$ when M_1 is large. In the practical applications below, we therefore use instead the expected value $\mathbb{E}_{(t,x) \sim \mu_1(Q)} [\mathcal{H}'(t, x, \theta, \omega)]$ to verify that the early termination conditions in Algorithm 6 are satisfied, as suggested in Duan et al. (2023). This appears to provide fast and very accurate results numerically, as investigated in Section 6.5 and 6.6. As in Algorithm 5, the loss function $\mathcal{L}_2'$ given in (6.3.7) can also be used in place of (6.3.12) for speeding up convergence in some empirical cases, even if it does not ensure to find a global minimum u^* of the Hamiltonian. Moreover, Assumption 6.2 allows to considerably simplify the exposition of the proof of Theorem 6.3. However, after closer inspection, it is easy to see that Assumption 6.2 is not necessary and can be dropped from the proof straightforwardly, leading to the following Corollary 6.4. The weights $\theta_N^{(k)}$ (resp. $\omega_N^{(k)}$) now denote the NN value (resp. control) weights at step k for a feedforward NN with one hidden layer of N neurons.

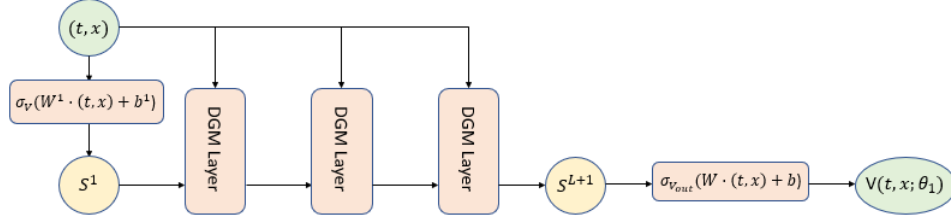
Corollary 6.4. *Let the assumptions of Theorem 6.3 still hold except for Assumption 6.2. Then, $\forall \varepsilon > 0, \exists k^*, N^* : k > k^*, N_1 > N^*, N_2 > N^*$, one has*

$$\sup_{y \in Q} |V(y; \theta_{N_1}^{(k)}) - V(y)| < \varepsilon \quad \text{and} \quad \sup_{y \in Q} |u(y; \omega_{N_2}^{(k)}) - u^*(y)| < \varepsilon.$$

Proof. The proof of Theorem 6.3 still holds using Lemma 6.1 when taking a sufficiently large number of neurons N_1, N_2 with $\theta_{N_1}^{(k)}, \omega_{N_2}^{(k)}$ and by introducing explicitly NN approximation errors ε . See Theorems 2–4 of Abu-Khalaf and Lewis (2005).

6.4 Neural network implementation

Instead of the shallow NN of Theorem 6.3, we choose in practice the DGM architecture introduced in Sirignano and Spiliopoulos (2018) for both the optimal

Figure 6.4.1: DGM architecture for the value function with $L = 3$ layers

control NN and the value NN, as it empirically speeds up the convergence to the true solutions. It is inspired by the LSTMs and Highway Networks introduced in Schmidhuber and Hochreiter (1997) and Srivastava et al. (2015), respectively. As pictured in Figure 6.4.1, each DGM layer takes as input the mini-batch inputs (t, x) (the set of randomly sampled time-space points from μ_1) and the output of the previous DGM layer. This DGM network then leads to a vector-valued output consisting of the NN approximation of the desired function $V(\cdot; \theta)$ or $u(\cdot; \omega)$ evaluated at the mini-batch points. Following a similar rationale as in LSTMs, each layer produces weights based on the previous layer, determining how much information is transferred to the next layer, which allows to effectively address the issue of vanishing gradients while remaining flexible enough to model complex functions. The number of parameters in each hidden layer of the DGM network is roughly eight times bigger compared with shallow NNs. Finally, we implement the Algorithm 6 using TensorFlow and Keras, which are software libraries in Python for deep learning. TensorFlow has reverse mode automatic differentiation which allows the exact calculation of derivatives for a broad range of functions (including the DGM neural network with respect to t , x , θ , and ω).

More precisely, each DGM layer in Fig. 6.4.1 for the value network is as follows

$$\begin{aligned}
 S^1 &= \sigma_V(W^1 \cdot (t, x) + b^1), \\
 Z^\ell &= \sigma_V(U^{z, \ell} \cdot (t, x) + W^{z, \ell} \cdot S^\ell + b^{z, \ell}), \quad \ell = 1, \dots, L, \\
 G^\ell &= \sigma_V(U^{g, \ell} \cdot (t, x) + W^{g, \ell} \cdot S^1 + b^{g, \ell}), \quad \ell = 1, \dots, L, \\
 R^\ell &= \sigma_V(U^{r, \ell} \cdot (t, x) + W^{r, \ell} \cdot S^\ell + b^{r, \ell}), \quad \ell = 1, \dots, L, \\
 H^\ell &= \sigma_V(U^{h, \ell} \cdot (t, x) + W^{h, \ell} \cdot (S^\ell \odot R^\ell) + b^{h, \ell}), \quad \ell = 1, \dots, L, \\
 S^{\ell+1} &= (1 - G^\ell) \odot H^\ell + Z^\ell \odot S^\ell, \quad \ell = 1, \dots, L, \\
 V(t, x; \theta) &= \sigma_{V_{out}}(W \cdot S^{L+1} + b),
 \end{aligned}$$

where the number of hidden layers is $L + 1$, and $\odot$ denotes element-wise multiplication. The parameters are

$$\theta = \left\{ W^1, b^1, (U^{z,\ell}, W^{z,\ell}, b^{z,\ell})_{\ell=1}^L, (U^{g,\ell}, W^{g,\ell}, b^{g,\ell})_{\ell=1}^L, (U^{r,\ell}, W^{r,\ell}, b^{r,\ell})_{\ell=1}^L, (U^{h,\ell}, W^{h,\ell}, b^{h,\ell})_{\ell=1}^L, W, b \right\}.$$

The number of units in each layer is N , $\sigma_V : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a twice-differentiable element-wise nonlinearity and $\sigma_{V_{out}} : \mathbb{R}^N \rightarrow \mathbb{R}^+$ is the output activation function (assumed to be non-negative and twice-differentiable). The control NN is designed following the same DGM architecture with $\sigma_{u_{out}} : \mathbb{R}^N \rightarrow \mathbb{R}^m$ locally bounded. In the examples below, we consider $L = 3$ layers each one with $N = 50$ neurons. Unless otherwise stated, we take $M_1 = M_2 = 5000$ leading to a sample size $M = 10\,000$, a batch size $b = 128$ and a maximum number of epochs $k_{\max} = 500$. The network's parameters are updated using the Adaptive Moment Estimation with constant learning rates η_0, η_1, η_2 equal to 0.005. Weights and biases are initialized using the Glorot uniform initializer. Finally, the value activation function σ_V is the tanh and $\sigma_{V_{out}}$ is the softplus function, as justified by Theorem 6.3. The control activation function σ_u is again the tanh and $\sigma_{u_{out}}$ the linear activation function.

6.5 A comparative example

We now compare the performance of our GPI-PINN Algorithm 6 on an example for which we can compute the solution easily. We consider the setting of Example 1 with a classical linear-quadratic Gaussian control problem in n dimensions with $Q = [0, T) \times \mathbb{R}^n$. We assume that the dimension m of the control is also n with $\mathbb{U} = \mathbb{R}^n$ and that $(W_t)_{t \in [0, T]}$ is a n -dimensional $\mathbb{F}$ -Brownian motion ($d = n$). We take as drift $f(t, x) = 2\sqrt{\lambda}I_n$ and diffusion coefficient $\sigma(t, x) = \sqrt{2}I_n$, where λ a positive constant representing the strength of the control. Finally, we suppose $\mathbf{R} = I_n$, $h(t, x) = 0$ and $\phi(x) = \ln((1 + \|x\|^2)/2)$, so that

$$V^u(t, x) = \mathbb{E} \left[\int_t^T \|u(s, X_s^u)\|^2 ds + \ln((1 + \|X_T^u\|^2)/2) \middle| X_t^u = x \right], \quad (6.5.1)$$

where X^u is the controlled process governed by

$$dX_s^u = 2\sqrt{\lambda}u_s ds + \sqrt{2}dW_s, \quad t \leq s \leq T, \quad X_t^u = x.$$

The cost-to-go (6.5.1) is indeed non-negative for any control $u \in \mathcal{A}$. The associated HJB equation for the value function $V(t, x) = \inf_{u \in \mathcal{A}} V^u(t, x)$ writes

$$\begin{aligned} 0 &= \partial_t V(t, x) + \text{Tr}[V_{xx}(t, x)] + \inf_{u \in \mathbb{U}} \left\{ \|u\|^2 + 2\sqrt{\lambda}[V_x(t, x)]^\top u \right\} \\ \Leftrightarrow 0 &= \partial_t V(t, x) + \sum_{i=1}^n V_{x_i x_i}(t, x) + \inf_{u \in \mathbb{U}} \left\{ \sum_{i=1}^n (u^{(i)})^2 + 2\sqrt{\lambda}V_{x_i}(t, x)u^{(i)} \right\}, \end{aligned} \quad (6.5.2)$$

with terminal condition $V(T, x) = \phi(x)$. Hence, the Hamiltonian is given by

$$\mathcal{H}(t, x, V_x, V_{xx}, u) = \|u\|^2 + 2\sqrt{\lambda} [V_x(t, x)]^\top u + \text{Tr} [V_{xx}(t, x)],$$

From Han et al. (2017), it holds that $V : \bar{Q} \rightarrow \mathbb{R}$ is $C^{1,2}(Q) \cap C(\bar{Q})$ and for $(t, x) \in \bar{Q}$ that

$$V(t, x) = -\frac{1}{\lambda} \ln \left(\mathbb{E} \left[\exp \left(-\lambda \phi(x + \sqrt{2} W_{T-t}) \right) \right] \right). \quad (6.5.3)$$

This representation will hence enable the use of Monte-Carlo (MC) simulations so as to compare our approximated solution from Algorithm 6 with the true solution (6.5.3). Since V is smooth, the optimal control is given by

$$u^*(t, x) = -\sqrt{\lambda} V_x(t, x), \quad (6.5.4)$$

for all $(t, x) \in Q$, where $V_x = (V_{x_1}, \dots, V_{x_n})^\top$ and with

$$V_{x_i}(t, x) = \frac{2}{\mathbb{E} [\exp(-\lambda \phi(x + \sqrt{2} W_{T-t}))]} \mathbb{E} \left[\frac{\exp(-\lambda \phi(x + \sqrt{2} W_{T-t})) (x_i + \sqrt{2} W_{T-t}^{(i)})}{1 + \|x + \sqrt{2} W_{T-t}\|^2} \right],$$

for $i = 1, \dots, n$. Hence, the solution of the n -dimensional optimal control u^* can also be obtained using MC samples. Moreover, as highlighted in Bachouch et al. (2022), the frequencies with which various states appear in the sample should be roughly proportional to the probabilities of their occurrence under an optimal policy. As one can guess from the representation of $V^u(t, x)$ in equation (6.5.1), it is probably optimal to drive the process X^u around 0. Hence, since we have $\mu_1(t, x) = \mu_1(x|t)\mu_1(t)$, we assume $\mu_1(t) \sim \mathcal{U}_{[0, T]}$ and $\mu_1(x|t) \sim \sqrt{t} \mathcal{N}_n(0, I_n)$ to simulate the values of x and t , with $\mathcal{U}$ a uniform law and $\mathcal{N}_n$ a n -dimensional Gaussian law. Similarly, $\mu_2(x) \sim \sqrt{T} \mathcal{N}_n(0, I_n)$. This leads to the following plots of the value function and optimal control comparing the numerical results from the representation (6.5.3)-(6.5.4) with 10^6 MC samples to the results derived from Algorithm 6, for different values of (t, x) with $T = 1$ and $\lambda = 1$. We consider a 10-dimensional problem ($n = 10$). Note that only the first component $u_1^*(t, x)$ of this 10-dimensional optimal control is depicted in the right plots of Figures 6.5.1-6.5.2. Moreover, the orange shaded area in both figures represents the standard deviation of the numerical solution computed on 10 independent runs of Algorithm 6.

We obtain highly accurate results using our GPI-PINN Algorithm across all time and space, for both the value NN and the optimal control NN. In the first Figure 6.5.1, we however observe a higher standard deviation for larger absolute values of x_1 (close to 1.5) since we are sampling training observations from $\mu_1(t, x)$ and hence, the majority (around 85%) of observations from the training data belongs to $[-1, 1]$. Therefore, only few observations are available for training the NNs at these larger absolute values of x_1 , which makes the NN outputs less precise. Similarly, the standard deviation in the second Figure 6.5.2 is increasing with t

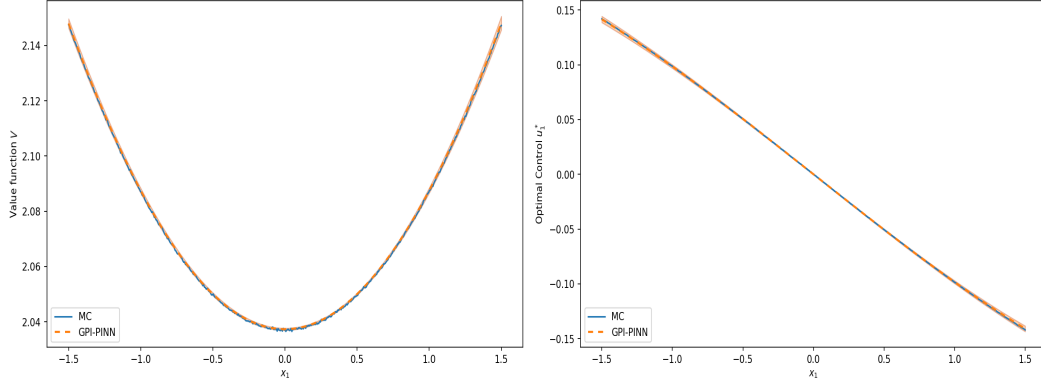


Figure 6.5.1: Value function $V(t, x)$ (left) and first component of the optimal control $u_1^*(t, x)$ (right) at $t = 0$ for $x = (x_1, \mathbf{0}_9)$ with $x_1 \in [-1.5, 1.5]$. In orange dotted line, numerical results from Algorithm 6 with $\pm$ one standard deviation in orange shaded area ($k_{\max} = 500$, $M = 20\,000$). In blue line, MC average of the solution (6.5.3)-(6.5.4).

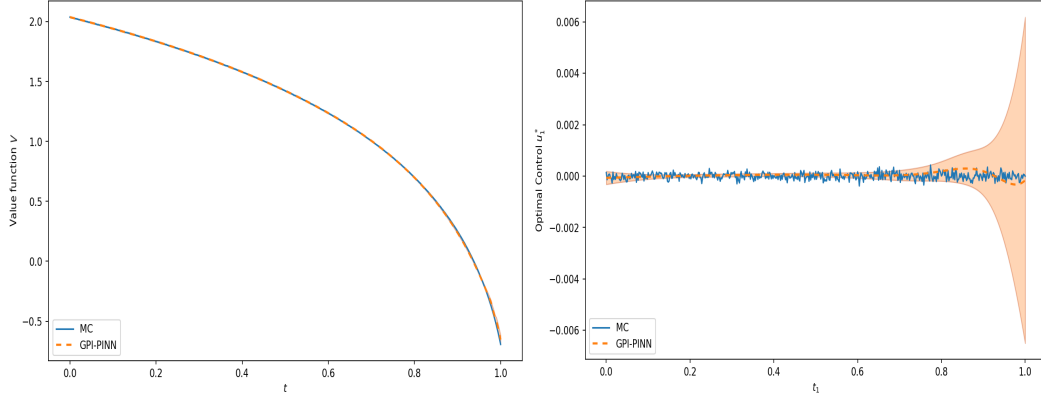


Figure 6.5.2: Optimal value function $V(t, x)$ (left) and first component of the optimal control $u_1^*(t, x)$ (right) at $x = \mathbf{0}_{10}$ with $t \in [0, 1]$. In orange dotted line, numerical results from Algorithm 6 with $\pm$ one standard deviation in orange shaded area ($k_{\max} = 500$, $M = 20\,000$). In blue line, MC average of the solution (6.5.3)-(6.5.4) with 10^6 samples.

since $\mu_1(x|t) \sim \sqrt{t}\mathcal{N}_n(0, I_n)$ and hence, the higher the time t , the higher the variance of the conditional Gaussian distribution and the less the generated observations are close to the true value 0. Finally, we obtain the following numerical results in Tables 6.5.1 and 6.5.2 from the GPI-PINN approximation of $V(t, x)$ and $u_1^*(t, x)$ for $t = 0$ and $x = (0.5, \mathbf{0}_9)$. Again, $u_1^*(t, x)$ is the first component of

the 10-dimensional optimal control $u^*(t, x)$. The standard deviations and means in both tables below are also obtained from 10 independent runs of Algorithm 6. The Mean Absolute Error (MAE) is computed between the mean of the value (resp. control) NN on these 10 independent runs and the mean of the 10^6 MC samples of representation (6.5.3) (resp. (6.5.4)). Finally, the loss $\mathcal{L}$ is defined as $\mathcal{L} := \frac{1}{2} \left(\widehat{\mathcal{L}}_1(\theta^{(k_{\max})}, \omega^{(k_{\max})}) + \widehat{\mathcal{L}}_2(\theta^{(k_{\max})}, \omega^{(k_{\max})}) \right)$, based on (6.3.8)-(6.3.9).

Table 6.5.1: Numerical results from the GPI-PINN Algorithm 6 for $V(0, (0.5, \mathbf{0}_9))$ and $u_1^*(0, (0.5, \mathbf{0}_9))$ with different sample size M and for a number of epochs $k_{\max} = 500$.

Sample size	Mean V	Std. V	Mean u_1^*	Std. u_1^*	MAE $_V$	MAE $_{u_1^*}$	Mean $\mathcal{L}$	Std. $\mathcal{L}$	Runtime(s)
100	2.0621	0.01687	-0.07787	0.00739	0.06752	0.02738	-6.10	2.0761	769
1000	1.9956	0.0109	-0.04163	0.00675	0.05407	0.00887	-55.58	2.5810	1674
5000	2.0474	0.00325	-0.04842	0.00304	0.00304	0.00295	-80.13	0.6722	7168
10 000	2.0503	0.00162	-0.05111	0.00078	0.00131	0.00088	-81.40	0.2342	11543
20 000	2.0499	0.00031	-0.05055	0.00059	0.00021	0.00005	-81.76	0.1946	29454

Table 6.5.2: Numerical results from the GPI-PINN Algorithm 6 for $V(0, (0.5, \mathbf{0}_9))$ and $u_1^*(0, (0.5, \mathbf{0}_9))$ with various number of epochs $k_{\max}$ and for a fixed sample size $M = 10\,000$.

Epochs	Mean V	Std. V	Mean u_1^*	Std. u_1^*	MAE $_V$	MAE $_{u_1^*}$	Mean $\mathcal{L}$	Std. $\mathcal{L}$	Runtime(s)
0	0.0908	0.18160	0.07019	0.11049	1.95890	0.14528	16 306 636	1 980 892	
50	1.9921	0.08227	-0.04185	0.00708	0.05754	0.00865	-49.6136	3.9522	2675
100	2.0213	0.01384	-0.04290	0.00623	0.02838	0.00759	-67.1391	3.2357	5501
500	2.0503	0.00162	-0.05111	0.00078	0.00131	0.00088	-81.3961	0.2342	11543
1000	2.0503	0.00043	-0.05058	0.00045	0.00061	0.00038	-81.8925	0.2017	32062

As expected, we clearly see that the higher the training size and the larger the number of epochs in the training procedure, the more accurate and precise are the outputs of the GPI-PINN Algorithm 6 for both the value function and the optimal control.

Compared to BSDE methods for semilinear PDEs, PINNs (and hence also the GPI-PINN algorithm) typically exhibit a slower convergence speed and a reduced accuracy, as emphasized in the review paper of Blechschmidt and Ernst (2021). However, as mentioned in Section 6.2.1, the deep BSDE method of Han et al. (2017) only provide local solutions at a single point in time $t_n \in [0, T]$. More precisely, for a temporal discretization of $[0, T]$ with N subintervals $0 = t_0 < t_1 < \dots < t_N = T$, it then needs to train N neural networks at the same time. Note that the deep

splitting method of Beck et al. (2021) trains one network after the other and can therefore handle larger problems since it decomposes them into smaller computational tasks which can be solved successively for $(t_n)_{0 \leq n \leq N}$. In contrast, the advantage of PINNs is that they do not require this discretized time grid and can solve the PDE for any time $t \in [0, T]$ and $x \in \mathbb{X}$ using a single neural network $V_\theta : [0, T] \times \mathbb{X} \rightarrow \mathbb{R}$ (and a second NN is then needed for the optimal control in the GPI-PINN algorithm). Moreover, to the best of our knowledge, these deep BSDE schemes do not include the possibility of having only an implicit representation for the optimal control since they cannot handle in this case the minimization part of the HJB PDE, see *e.g.* equation (6.2.19).

Finally, we want to compare the accuracy and runtime between standard PINNs and our GPI-PINN Algorithm 6, when an explicit control is available. This will allow us to assess the price not to exploit this explicit control when available. From equation (6.5.4), we have that the optimal control in Example 6.5 admits an explicit representation and we can then rewrite the HJB PDE (6.5.2) explicitly

$$0 = \partial_t V(t, x) + \sum_{i=1}^n (V_{x_i x_i}(t, x) - \lambda V_{x_i}^2(t, x)).$$

This explicit PDE enables the use of the (DGM) PINN algorithm of Sirignano and Spiliopoulos (2018), see Section 6.3.1. We therefore obtain in Figure 6.5.3 two plots comparing the accuracy in terms of MAE_V as well as the runtime between the DGM and the GPI-PINN algorithm in function of the training size and the number of epochs.

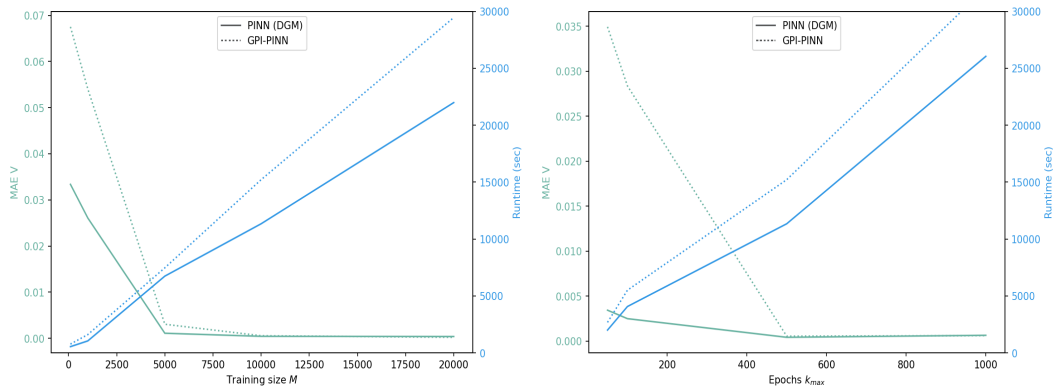


Figure 6.5.3: Comparison of the MAE_V (green) and the runtime in seconds (blue) between the DGM algorithm (solid line) and the GPI-PINN algorithm (dashed line) in function of the training size (left) and the number of epochs (right).

Clearly, the MAE_V for the value function V between the two methods converges when the number of epochs or the training size increases. In contrast, the higher the number of epochs or the training size, the more exacerbated is the difference in runtime between the two algorithms. Hence, an equivalent accuracy can be attained in the GPI-PINN algorithm but at the cost of a slower runtime, or the same runtime can be imposed but at the cost of a higher MAE_V . A good compromise would be to take $M = 10\,000$ and $k_{\max} = 500$.

6.6 A practical example in mathematical finance

As seen in the previous chapter, a classical problem in finance is how an agent can sell a large amount of shares and yet minimize adverse price movements generated by her own trades. The agent must then take into account the balance between price impact (trade quicker) and price risk (take longer to complete all trades). Indeed, if an agent trades quickly, then her orders will walk through the buy side of the limit order book (LOB) and she will obtain lower prices for her orders, leading to a *temporary* price impact. Even if she splits each order into smaller trades so that each one does not walk the book, then other traders will notice an excess of sell orders and adjust their quotes inducing again a negative price impact, which is in this case *permanent*. On the other hand, if she trades slowly so as to avoid this price impact, then she will face the uncertainty of future prices, which could have dropped to a less desirable level by the time the agent executes the whole order. The aim of optimal liquidation is hence to propose an optimal trading strategy such as to balance these two factors (price impact vs. market risk), based on a coherent price impact model reproducing the underlying asset price and the impact of the agent's transactions on this price. Following the setting of Chapter 5, we denote by T the finite trading horizon required for liquidating the shares. The trading rate³ $\nu = (\nu_t)_{0 \leq t \leq T}$ is the (positive) speed at which the agent is liquidating shares using market orders and which can be seen as the volume sold on each infinitesimal step of time. Q^ν is the agent's inventory of remaining shares to liquidate, affected by how fast she trades and given by $dQ_t^\nu = -\nu_t dt$ with $Q_0^\nu = Q$. S^ν is the midprice of the asset, affected by the speed of trading ν and given by

$$dS_t^\nu = -g(\nu_t)dt + \sigma dW_t, \quad (6.6.1)$$

where $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes the permanent price impact of the agent's trading and $W = (W_t)_{0 \leq t \leq T}$ is a standard $\mathbb{F}$ -Brownian motion. The execution price $\hat{S}^\nu$ is then the price at which the agent can effectively sell the asset (also affected by the trading rate) :

$$\hat{S}_t^\nu = S_t^\nu - f(\nu_t), \quad (6.6.2)$$

³Standard notation for the trading rate replaces the control u used until now by ν .

where $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes the temporary price impact that the agent's trading has on the execution price. Finally, X^ν is the agent's cash process resulting from her execution strategy and defined by $X_t^\nu = \int_0^t \hat{S}_s^\nu \nu_s ds = \int_0^t (S_s^\nu - f(\nu_s)) \nu_s ds$. As classically proposed in the literature (see Chapter 5 and Cartea et al. (2014)), we assume that the agent is willing to optimize the following performance criterion

$$V^\nu(t, x, S, q) = \mathbb{E}_{t,x,S,q} \left[\underbrace{X_T^\nu}_{\text{Terminal Wealth}} + \underbrace{Q_T^\nu (S_T^\nu - \pi Q_T^\nu)}_{\text{Terminal Execution}} - \underbrace{\phi \int_t^T (Q_s^\nu)^2 ds}_{\text{Inventory Penalty}} \right], \quad (6.6.3)$$

with respect to the trading rate ν , and where $\mathbb{E}_{t,x,S,q}[\cdot]$ denotes the expectation conditional to $X_t^\nu = x$, $S_t^\nu = S$, $Q_t^\nu = q$. The agent thus wants to optimize her expected wealth at maturity minus some penalties. If the agent's strategy reaches the terminal date T with inventory left, then she must execute an order to liquidate the remaining shares for a total revenue of $Q_T^\nu (S_T^\nu - \pi Q_T^\nu)$, where $\pi \geq 0$ is the terminal penalty parameter. We also introduce a running inventory penalty of the form $\phi \int_t^T (Q_s^\nu)^2 ds$ with $\phi \geq 0$. This running inventory penalty is not a financial cost to the agent's strategy but instead allows to incorporate the agent's urgency for executing the trade (risk-aversion penalty). The value function is then

$$V(t, x, S, q) = \sup_{\nu \in \mathcal{A}} V^\nu(t, x, S, q), \quad (6.6.4)$$

where $\mathcal{A}$ is the set of admissible strategies which consist in $\mathbb{F}$ -predictable, non-negative and bounded from above trading rates ν . Standard results in dynamic programming, see Cartea et al. (2015), imply that the value function V is a classical solution of the following HJB PDE for all $(t, S, q) \in [0, T) \times \mathbb{R} \times (0, Q]$:

$$0 = \left(\partial_t + \frac{1}{2} \sigma^2 \partial_{SS} \right) V - \phi q^2 + \sup_{\nu \in \mathcal{A}} \{ (\nu(S - f(\nu)) \partial_x - g(\nu) \partial_S - \nu \partial_q) V \}, \quad (6.6.5)$$

with terminal condition $V(T, x, S, q) = x + q(S - \pi q)$ and boundary condition $V(t, x, S, 0) = 0$ since there are no more shares to liquidate and since the agent cannot buy back shares (non-negative trading rate). Using integration by parts, it is easy to show that

$$\begin{aligned} V(t, x, S, q) &= x + Sq + \sup_{\nu \in \mathcal{A}} \mathbb{E}_{t,q} \left[- \int_t^T f(\nu_s) \nu_s ds - \int_t^T g(\nu_s) Q_s^\nu ds - \pi (Q_T^\nu)^2 - \phi \int_t^T (Q_s^\nu)^2 ds \right] \\ &= x + Sq - \inf_{\nu \in \mathcal{A}} \mathbb{E}_{t,q} \left[\int_t^T f(\nu_s) \nu_s ds + \int_t^T g(\nu_s) Q_s^\nu ds + \pi (Q_T^\nu)^2 + \phi \int_t^T (Q_s^\nu)^2 ds \right] \\ &= x + Sq - \mathcal{V}(t, q), \end{aligned} \quad (6.6.6)$$

where $\mathcal{V}(t, q)$ is again a value function but which now states for the penalty stemming from optimally liquidating the remaining of the shares. It is clearly non-negative and it is again a classical solution of the following HJB PDE for all

$(t, q) \in [0, T) \times (0, Q]$:

$$0 = \partial_t \mathcal{V}(t, q) + \phi q^2 + \inf_{\nu \in \mathcal{A}} \{ \nu f(\nu) + g(\nu)q - \nu \partial_q \mathcal{V}(t, q) \}, \quad (6.6.7)$$

with $\mathcal{V}(T, q) = \pi q^2$ and $\mathcal{V}(t, 0) = 0$. Let us assume that $\nu f(\nu)$ and $g(\nu)$ are convex functions in ν . Almgren et al. (2005), Gatheral (2010) and Cartea et al. (2014) have shown that a power-law function with exponent less than one is the most suitable choice for the temporary price impact $f(\nu) = \kappa \nu^\gamma$ with $\kappa \geq 0, \gamma \in [0, 1]$. The permanent impact function is also here assumed to be a power-law function of the form $g(\nu) = b \nu^\eta$ with $b \geq 0, \eta \geq 1$, which is consistent with Figure 1 of Cartea and Jaimungal (2016) (see also Chapter 5). This allows us to generalize the classical models of Almgren and Chriss (2001) and Almgren (2003) by considering both nonlinear temporary and permanent impact functions. We finally obtain that the value function $\mathcal{V}$ must satisfy for all $(t, q) \in [0, T) \times (0, Q]$:

$$0 = \partial_t \mathcal{V}(t, q) + \phi q^2 + \inf_{\nu \in \mathcal{A}} \{ \kappa \nu^{\gamma+1} + b \nu^\eta q - \nu \partial_q \mathcal{V}(t, q) \}, \quad (6.6.8)$$

with $\mathcal{V}(T, q) = \pi q^2$ and $\mathcal{V}(t, 0) = 0$. For general exponents $\gamma \in [0, 1]$ and $\eta \geq 1$, we directly see from (6.6.8) that the first order condition, $\kappa(\gamma+1)\nu^\gamma + b\eta\nu^{\eta-1}q - \partial_q \mathcal{V}(t, q) = 0$, does not lead to an explicit solution for the optimal control ν^* , requiring the use of the GPI-PINN Algorithm 6. In the classical Almgren-Chriss model $\gamma = 1$ and $\eta = 1$ studied in Almgren and Chriss (2001) and in Cartea et al. (2014), it is however shown that

$$\mathcal{V}(t, q) = q^2 \left(-\frac{1}{2}b + \chi(t) \right), \quad (6.6.9)$$

with $\chi(t) = \sqrt{\kappa\phi} \frac{1+\zeta e^{2\rho(T-t)}}{1-\zeta e^{2\rho(T-t)}}$, $\rho = \sqrt{\phi/\kappa}$ and $\zeta = \frac{(\pi-0.5b+\sqrt{\kappa\phi})}{(\pi-0.5b-\sqrt{\kappa\phi})}$. Moreover, this leads to

$$\nu^*(t, q) = \gamma \frac{\zeta e^{\rho(T-t)} + e^{-\rho(T-t)}}{\zeta e^{\rho(T-t)} - e^{-\rho(T-t)}} q. \quad (6.6.10)$$

We can therefore compare these analytical solutions of the value function (6.6.9) and optimal control (6.6.10) with $\gamma = 1$ and $\eta = 1$ to the numerical results output by our GPI-PINN Algorithm 6, as depicted in Figures 6.6.1 and 6.6.2. In this one-dimensional optimal liquidation setting, we clearly see that the GPI-PINN algorithm manages to perfectly approximate both the value function and the optimal control across all time and space. The third Figure 6.6.3 compares the analytical solution in the case $\gamma = \eta = 1$ to the implicit case $\gamma = 0.7, \eta = 1.5$ solved using the GPI-PINN Algorithm 6, allowing to assess the impact of the parameters γ and η on the optimal strategies.

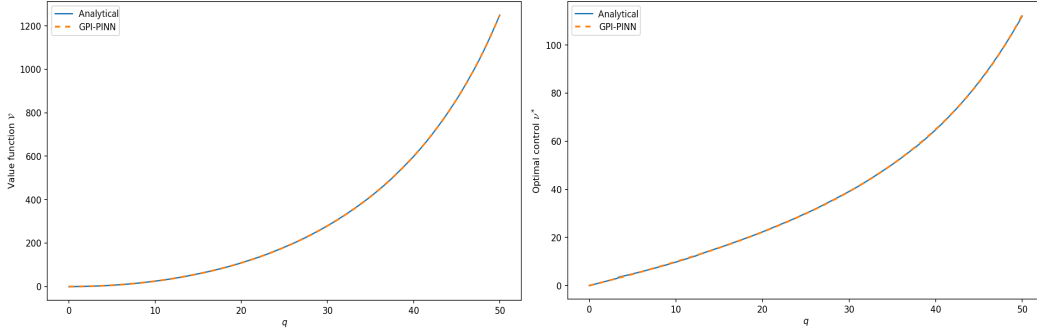


Figure 6.6.1: Value function $\mathcal{V}(t, q)$ and optimal control $\nu^*(t, q)$ at $t = 0$ for $q \in [0, 50]$ with $\gamma = 1, \eta = 1, T = 1, \kappa = 0.2, b = 0.1, \pi = 0.5, \phi = 0.1$. In orange dotted line, numerical solution from Algorithm 6 ($k_{\max} = 500, M = 20\,000$). In blue line, analytical solutions (6.6.9)-(6.6.10).

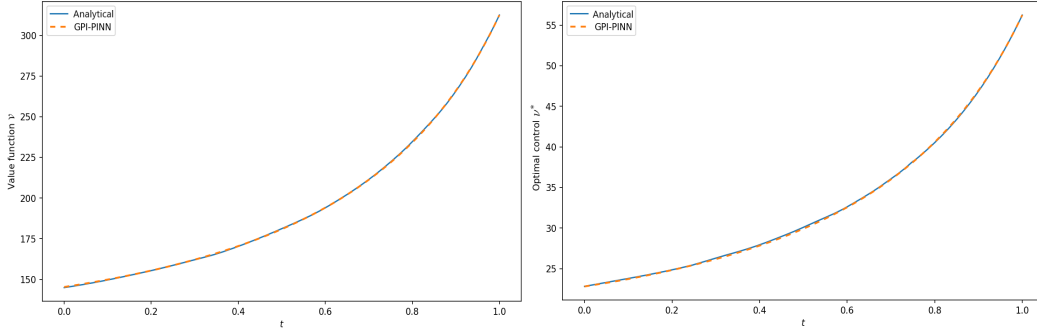


Figure 6.6.2: Value function $\mathcal{V}(t, q)$ and optimal control $\nu^*(t, q)$ at $q = 25$ for $t \in [0, 1]$, with $\gamma = 1, \eta = 1, T = 1, \kappa = 0.2, b = 0.1, \pi = 0.5, \phi = 0.1$. In orange dotted line, numerical solution from Algorithm 6 ($k_{\max} = 500, M = 20\,000$). In blue line, analytical solutions from equations (6.6.9)-(6.6.10).

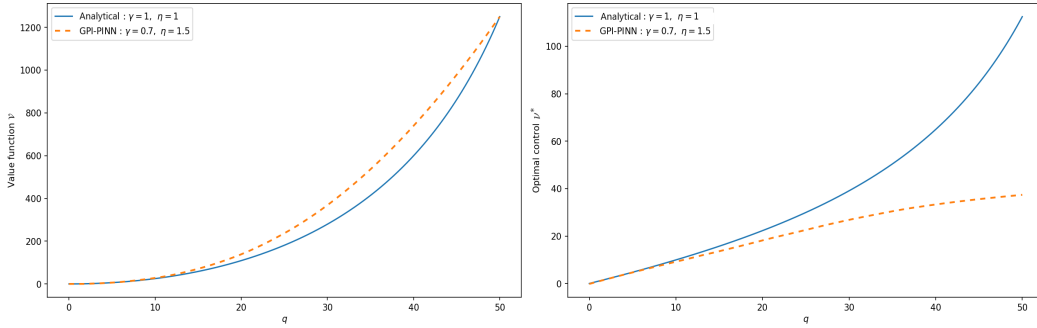


Figure 6.6.3: Value function $\mathcal{V}(t, q)$ and optimal control $\nu^*(t, q)$ at $t = 0$ for $q \in [0, 50]$ with $T = 1, \kappa = 0.2, b = 0.1, \pi = 0.5, \phi = 0.1$. In orange dotted line, numerical solution from Algorithm 6 with $\gamma = 0.7, \eta = 1.5$ ($k_{\max} = 500, M = 20\,000$). In blue line, analytical solution from (6.6.9)-(6.6.10) with $\gamma = 1, \eta = 1$.

6.6.1 Multi-asset optimal liquidation

We now consider a multi-asset optimal liquidation setting since the GPI-PINN algorithm has been especially devised for solving high-dimensional (implicit) optimal control problems. The studied financial portfolio now consists in n assets which are co-integrated in the sense that each asset's trading permanently affects the price at which all the other assets are executed. Indeed, market sell orders in one stock also exert in this setting a permanent downward pressure on the price of co-integrated assets. As explained in Cartea et al. (2019), this cross-effect is partly caused by trading algorithms that take positions based on the co-movements of assets. We also include correlations between each pair of assets in the portfolio. Therefore, the above one-dimensional optimal liquidation model is enlarged in the following way. The trading rate in the i^{th} stock is denoted $\nu^{(i)} = (\nu_t^{(i)})_{0 \leq t \leq T}$ and the related remaining inventory satisfies $dQ_t^{\nu,i} = -\nu_t^{(i)} dt$, for all $i = 1, \dots, n$. We also denote $\boldsymbol{\nu}_t = (\nu_t^{(1)}, \dots, \nu_t^{(n)})^\top$ and $\mathbf{Q}_t^\nu = (Q_t^{\nu,1}, \dots, Q_t^{\nu,n})^\top$. The midprice of the i^{th} stock is the i^{th} component of $\mathbf{S}_t^\nu = (S_t^{\nu,i}, \dots, S_t^{\nu,n})^\top$ and follows

$$dS_t^{\nu,i} = -g^{(i)}(\boldsymbol{\nu}_t)dt + \sigma_i dW_t^{(i)}, \quad (6.6.11)$$

where the function $g^{(i)} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ describes the permanent impact of each asset's trading on the midprice of the i^{th} asset. We also have that $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)^\top$ is the Cholesky decomposition of the asset prices' correlation matrix $\tilde{\boldsymbol{\Sigma}}$ (i.e., $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$) and $\mathbf{W} = (\mathbf{W}_t)_{0 \leq t \leq T}$ is a n -dimensional $\mathbb{F}$ -Brownian motion. The execution price of the i^{th} asset then follows

$$\hat{S}_t^{\nu,i} = S_t^{\nu,i} - f^{(i)}(\nu_t^{(i)}),$$

where the function $f^{(i)} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ describes the temporary impact of trading in stock i , which only affects in this case the execution price of the corresponding asset i . Indeed, as the temporary price impact for asset i comes from the market orders walking the LOB of this specific asset, it does not induce cross-effects from other assets. The wealth of the agent from liquidating the portfolio is given by $X_t^\nu = \sum_{i=1}^n \int_0^t \hat{S}_s^{\nu,i} \nu_s^{(i)} ds$. As in the one-dimensional setting (6.6.3), we want to optimize the following performance criterion with respect to the vector of trading rates $\boldsymbol{\nu}$:

$$V^\nu(t, x, \mathbf{S}, \mathbf{q}) = \mathbb{E}_{t,x,\mathbf{S},\mathbf{q}} \left[X_T^\nu + (\mathbf{Q}_T^\nu)^\top \mathbf{S}_T^\nu - \pi(\mathbf{Q}_T^\nu)^\top \mathbf{Q}_T^\nu - \phi \int_t^T (\mathbf{Q}_s^\nu)^\top \tilde{\boldsymbol{\Sigma}} \mathbf{Q}_s^\nu ds \right] \quad (6.6.12)$$

conditionally to $X_t^\nu = x$, $\mathbf{S}_t^\nu = \mathbf{S} := (S_1, \dots, S_n)^\top$ and $\mathbf{Q}_t^\nu = \mathbf{q} := (q_1, \dots, q_n)^\top$. The value function is then $V(t, x, \mathbf{S}, \mathbf{q}) = \sup_{\boldsymbol{\nu} \in \mathcal{A}} V^\nu(t, x, \mathbf{S}, \mathbf{q})$ with $\mathcal{A}$ the set of admissible strategies consisting in non-negative, $\mathbb{F}$ -predictable processes such that $\int_0^T |\nu_s^{(i)}| ds < +\infty$, $\mathbb{P}$ -a.s. for each asset $i = 1, \dots, n$. Using integration by parts as

in (6.6.6), we can show that

$$\begin{aligned}
V(t, x, \mathbf{S}, \mathbf{q}) &= x + \mathbf{q}^\top \mathbf{S} + \sup_{\boldsymbol{\nu} \in \mathcal{A}} \mathbb{E}_{t, \mathbf{q}} \left[-\pi(\mathbf{Q}_T^\nu)^\top \mathbf{Q}_T^\nu - \phi \int_t^T (\mathbf{Q}_s^\nu)^\top \tilde{\boldsymbol{\Sigma}} \mathbf{Q}_s^\nu ds \right. \\
&\quad \left. - \sum_{i=1}^n \left(\int_t^T f^{(i)}(\nu_s^{(i)}) \nu_s^{(i)} ds + \int_t^T g^{(i)}(\nu_s) Q_s^{\nu, i} ds \right) \right] \\
&= x + \mathbf{q}^\top \mathbf{S} - \inf_{\boldsymbol{\nu} \in \mathcal{A}} \mathbb{E}_{t, \mathbf{q}} \left[\pi(\mathbf{Q}_T^\nu)^\top \mathbf{Q}_T^\nu + \phi \int_t^T (\mathbf{Q}_s^\nu)^\top \tilde{\boldsymbol{\Sigma}} \mathbf{Q}_s^\nu ds \right. \\
&\quad \left. + \sum_{i=1}^n \left(\int_t^T f^{(i)}(\nu_s^{(i)}) \nu_s^{(i)} ds + \int_t^T g^{(i)}(\nu_s) Q_s^{\nu, i} ds \right) \right] \\
&= x + \mathbf{q}^\top \mathbf{S} - \mathcal{V}(t, \mathbf{q}).
\end{aligned}$$

The dynamic programming principle suggests that the value function $\mathcal{V}$ is a classical solution to the HJB equation for all $(t, \mathbf{q}) \in [0, T) \times ([0, Q]^n \setminus \mathbf{0}_n)$:

$$0 = \partial_t \mathcal{V}(t, \mathbf{q}) + \phi \mathbf{q}^\top \tilde{\boldsymbol{\Sigma}} \mathbf{q} + \inf_{\boldsymbol{\nu} \in \mathcal{A}} \left\{ \sum_{i=1}^n \left(f^{(i)}(\nu^{(i)}) \nu^{(i)} + g^{(i)}(\boldsymbol{\nu}) q_i - \nu^{(i)} \partial_{q_i} \mathcal{V}(t, \mathbf{q}) \right) \right\}, \quad (6.6.13)$$

with terminal and boundary conditions $\mathcal{V}(T, \mathbf{q}) = \pi \mathbf{q}^\top \mathbf{q}$ and $\mathcal{V}(t, \mathbf{0}_n) = 0$. As motivated in the one-dimensional setting, we choose the following power-law functions for the permanent and temporary impact functions $g^{(i)}$ and $f^{(i)}$ with $\gamma_i \in [0, 1]$, $\kappa_i \geq 0$, $\eta_{i,j} \geq 1$, $b_{i,j} \geq 0$ for all $i, j = 1, \dots, n$:

$$g^{(i)}(\boldsymbol{\nu}) = \sum_{j=1}^n b_{i,j} \left(\nu^{(j)} \right)^{\eta_{i,j}} \quad \text{and} \quad f^{(i)}(\nu^{(i)}) = \kappa_i \left(\nu^{(i)} \right)^{\gamma_i}.$$

We indeed have with this parametrization that the permanent impact on the transaction price of asset i depends on all the assets $j = 1, \dots, n$ through their trading rate $\nu^{(j)}$, while its temporary impact only depends on the related asset $\nu^{(i)}$. Equation (6.6.13) hence becomes

$$0 = \partial_t \mathcal{V}(t, \mathbf{q}) + \phi \mathbf{q}^\top \tilde{\boldsymbol{\Sigma}} \mathbf{q} + \inf_{\boldsymbol{\nu} \in \mathcal{A}} \left\{ \sum_{i=1}^n \left(\kappa_i \left(\nu^{(i)} \right)^{\gamma_i+1} + \sum_{j=1}^n b_{i,j} \left(\nu^{(j)} \right)^{\eta_{i,j}} q_i - \nu^{(i)} \partial_{q_i} \mathcal{V}(t, \mathbf{q}) \right) \right\}. \quad (6.6.14)$$

First order condition on $\nu^{(i)}$ with $\eta_{i,j} \geq 1$, $\gamma_i \in [0, 1]$ leads for all $i = 1, \dots, n$ to

$$\kappa_i (\gamma_i + 1) \left(\nu^{(i)} \right)^{\gamma_i} + \sum_{j=1}^n b_{j,i} \eta_{j,i} \left(\nu^{(i)} \right)^{\eta_{j,i}-1} q_j - \partial_{q_i} \mathcal{V}(t, \mathbf{q}) = 0. \quad (6.6.15)$$

We see that if $\eta_{j,i} = 1$ for all j (*i.e.* linear permanent impacts), then the optimal

trading rate $\nu^{(i)*}$ can be obtained explicitly for all $i = 1, \dots, n$ by

$$\nu^{(i)*}(t, \mathbf{q}) = \left(\frac{-\sum_{j=1}^n b_{j,i} q_j + \partial_{q_i} \mathcal{V}(t, \mathbf{q})}{\kappa_i(\gamma_i + 1)} \right)^{1/\gamma_i}.$$

However, for general exponents $\eta_{j,i} \geq 1$, we again have from the first-order condition (6.6.15) that the optimal trading rate $\nu^{(i)*}$ for each asset $i = 1, \dots, n$ does not admit an explicit expression. We hence again face an implicit high-dimensional HJB equation (when n is large), requiring the use of the GPI-PINN Algorithm 6. We then obtain the following results with $n = 3$ (for readability purposes).

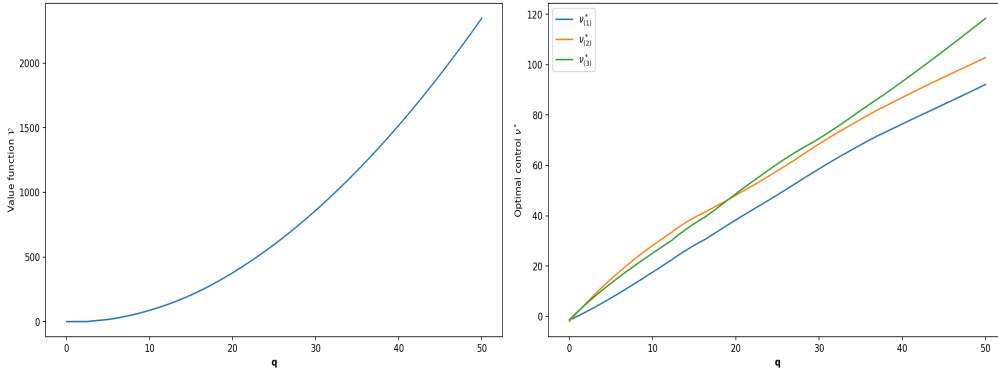


Figure 6.6.4: Optimal value function $\mathcal{V}(t, \mathbf{q})$ and optimal trading rate $\nu^*(t, \mathbf{q})$ for $\mathbf{q} \in [0, 50]^3$ and $t = 0.5$ with $n = 3$ assets from Algorithm 6 and the parameters below.

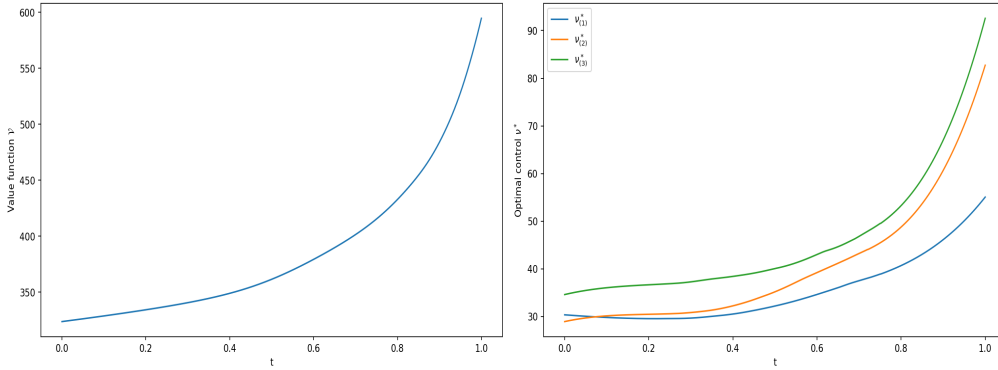


Figure 6.6.5: Optimal value function $\mathcal{V}(t, \mathbf{q})$ and optimal trading rate $\nu^*(t, \mathbf{q})$ for $t \in [0, 1]$ and $\mathbf{q} = (25, 25, 25)$ with $n = 3$ assets from Algorithm 6 and the parameters below.

The parameters used to derive these figures are as follows : $T = 1$, $\pi = 0.5$, $\phi = 0.2$, $\gamma = (0.95, 1, 0.92)$, $\kappa = (0.2, 0.1, 0.15)$, and

$$\eta = \begin{pmatrix} 1 & 1.05 & 1 \\ 1.07 & 1 & 1.1 \\ 1 & 1.11 & 1.02 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0.1 & 0.03 & 0.02 \\ 0.01 & 0.12 & 0 \\ 0.01 & 0.02 & 0.05 \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} 1 & 0.2 & 0.13 \\ 0.2 & 1 & 0.05 \\ 0.05 & 0.18 & 1 \end{pmatrix}.$$

6.7 Conclusion

Chapter 6 proposes a new successive deep-learning algorithm able to solve CT stochastic optimal control problems in high dimensions, even when the control does not admit an explicit representation. This numerical scheme is built from the combination of PINNs along with an Actor-Critic structure relying on the GPI procedure. By approximating the value function and the optimal control using two separate NNs, the GPI-PINN algorithm provides a global solution for stochastic optimal control problems across all time and space, which can be evaluated online fast. Thanks to an early termination scheme in the training process, we prove the uniform convergence of our algorithm toward the true solution of the value function and the optimal control. These convergence guarantees are however essentially qualitative since based on several (strong) assumptions. In particular, further research is needed to drop Assumption 6.1 from Theorem 6.3. Besides these theoretical guarantees, we also show empirically the efficacy and accuracy of the GPI-PINN scheme thanks to a comparative example for which the solution is known up to MC simulations. Finally, we provide a second practical example from mathematical finance by generalizing the well-known Almgren-Chriss model arising from optimal liquidation. More precisely, we introduce optimal liquidation strategies based on fully nonlinear temporary and permanent impact functions in a multidimensional setting with co-integrated assets, leading to an implicit optimal control problem in high dimensions. Once again, this example highlights the benefits of using the GPI-PINN algorithm to solve implicit high-dimensional CT stochastic optimal control problems and paves the way for various other applications in stochastic optimal control.

6.A Proof of Theorem 6.3

Recall first that the cost-to-go associated with the non-optimal Markov control u is defined as

$$V^u(t, x) = \mathbb{E}^{\mathbb{P}} \left[\phi(T, X_T^u) + \int_t^T \ell(\tau, X_\tau^u, u(\tau, X_\tau^u)) d\tau \middle| X_t^u = x \right]. \quad (6.A.1)$$

As in Oksendal (2013), we propose the change of variable $y = (t, x)$ and $Y_\tau^u = (\tau, X_\tau^u)$ so as to fall back to a time-invariant optimal control problem. Hence, we can rewrite

$$V^u(y) = \mathbb{E}_y^{\mathbb{P}} \left[\phi(Y_T^u) + \int_t^T \ell(Y_\tau^u, u(Y_\tau^u)) d\tau \right],$$

with $\mathbb{E}_y^{\mathbb{P}}$ the expectation conditional to $Y_t^u = y$ under the probability measure $\mathbb{P}$ and the dynamic

$$dY_\tau^u = (d\tau, f(Y_\tau^u, u(Y_\tau^u))d\tau + \sigma(Y_\tau^u, u(Y_\tau^u))dW_\tau^{\mathbb{P}}). \quad (6.A.2)$$

From the dynamic programming principle, we find with $h > 0$

$$V^u(y) = \mathbb{E}_y^{\mathbb{P}} \left[\int_t^{t+h} \ell(Y_\tau^u, u(Y_\tau^u)) d\tau + V^u(Y_{t+h}^u) \right]. \quad (6.A.3)$$

Since $V^u \in C^2(Q)$, using Dynkin's formula

$$\mathbb{E}_y^{\mathbb{P}} [V^u(Y_{t+h}^u)] = V^u(y) + \mathbb{E}_y^{\mathbb{P}} \left[\int_t^{t+h} (\mathcal{L}^u V^u)(Y_\tau^u) d\tau \right],$$

where $\mathcal{L}$ is the infinitesimal generator associated with the diffusion (6.A.2),

$$(\mathcal{L}^u \phi)(y) := \frac{\partial \phi}{\partial t}(y) + \sum_{i=1}^n f_i(y, u(y)) \frac{\partial \phi}{\partial x_i}(y) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\sigma \sigma^\top)_{ij}(y, u(y)) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(y). \quad (6.A.4)$$

In the limit $h \rightarrow 0$, we then find from the mean-value theorem for all $y \in Q$ the nonlinear Lyapunov equation

$$0 = \ell(y, u(y)) + (\mathcal{L}^u V^u)(y). \quad (6.A.5)$$

The notation (6.2.4) for the Hamiltonian $\mathcal{H}(t, x, V_x^u(t, x), V_{xx}^u(t, x), u(t, x))$ is simplified as $\mathcal{H}(y, V^u(y), u(y))$ and hence, (6.2.13) becomes $\forall y \in Q$,

$$\mathcal{H}'(y, V^u(y), u(y)) := \mathcal{H}(y, V^u(y), u(y)) + \frac{\partial V^u}{\partial t}(y) = \ell(y, u(y)) + (\mathcal{L}^u V^u)(y) = 0,$$

from the Lyapunov equation (6.A.5). When considering⁴ the NNs $V_{\theta^{(k)}}(y) := V(y; \theta^{(k)})$ and $u_{\omega^{(k)}}(y) := u(y; \omega^{(k)})$, we rewrite

$$\mathcal{H}'(y, \theta^{(k)}, \omega^{(k)}) := \mathcal{H}'(y, V_{\theta^{(k)}}(y), u_{\omega^{(k)}}(y)),$$

which is in general not equal to 0 since $V_{\theta^{(k)}}(y)$ is not governed by the control $u_{\omega^{(k)}}(y)$ (i.e. two separate NNs). First, note that since $\sigma_{u_{out}}(\cdot)$ is locally bounded, $\left(u(t, X_t^u; \omega)\right)_{0 \leq t \leq T} \in \mathcal{A}$ for all $\omega \in \mathbb{R}^p$. Then, since $\sigma_{V_{out}}(\cdot) \geq 0$, we have $V_{\theta}(y) \geq 0$ for all $y \in \bar{Q}$ and all $\theta \in \mathbb{R}^p$. Moreover, since $\sigma_V(\cdot)$ and $\sigma_{V_{out}}(\cdot)$ are twice differentiable, we have $V_{\theta}(y) \in C^2(\bar{Q})$. Finally, Assumption 6.2 allows us to ignore the NN approximation error ε when approximating $\phi(T, x)$ by $V_{\theta}(T, x)$. The loss function (6.3.4) then becomes

$$\begin{aligned} \mathcal{L}_1(\theta, \omega) &= \mathbb{E}_{y \sim \mu_1(Q)} \left[\mathcal{H}'(y, \theta, \omega)^2 \right] + \mathbb{E}_{x \sim \mu_2(\mathbb{X})} \left[(V_{\theta}(T, x) - \phi(T, x))^2 \right] \\ &= \mathbb{E}_{y \sim \mu_1(Q)} \left[\mathcal{H}'(y, \theta, \omega)^2 \right], \end{aligned} \quad (6.A.6)$$

and the loss (6.3.6) becomes

$$\mathcal{L}_2(\theta, \omega) = \mathbb{E}_{y \sim \mu_1(Q)} [\mathcal{H}'(y, \theta, \omega)].$$

According to (6.A.5), Lemma 6.1⁵ and Assumption 6.2, there exist $(\theta^{\dagger}, \omega^{\dagger})$ such that $\mathcal{H}'(y, \theta^{\dagger}, \omega^{\dagger}) = 0$ for all $y \in Q$, which implies

$$\min_{\theta, \omega} \mathcal{L}_1(\theta, \omega) = \min_{\theta, \omega} \mathbb{E}_{y \sim \mu_1(Q)} \left[\mathcal{H}'(y, \theta, \omega)^2 \right] = 0.$$

Since in Phase 1, Algorithm 6 updates θ and ω to continuously minimize $\mathcal{L}_1(\theta, \omega)$ using (6.3.10), the global minimum of $\mathcal{L}_1(\theta, \omega)$ can be attained in polynomial time. Therefore, before this global minimum, there exists $k^{\dagger} \in \mathbb{N}$ such that

$$\mathcal{H}'(y, \theta^{(k^{\dagger})}, \omega^{(k^{\dagger})}) \leq 0, \quad \forall y \in Q.$$

Then, Algorithm 6 enters Phase 2 and again from (6.A.5), we have

$$\min_{\theta} \mathcal{L}_1(\theta, \omega^{(k^{\dagger})}) = \min_{\theta} \mathbb{E}_{y \sim \mu_1(Q)} \left[\mathcal{H}'(y, \theta, \omega^{(k^{\dagger})})^2 \right] = 0.$$

From Assumption 6.1 and Lemma 6.1, we can always find $\theta^{(k^{\dagger}+1)}$ by continuously applying (6.3.11) such that

$$\mathcal{H}'(y, \theta^{(k^{\dagger})}, \omega^{(k^{\dagger})}) \leq \mathcal{H}'(y, \theta^{(k^{\dagger}+1)}, \omega^{(k^{\dagger})}) \leq 0, \quad \forall y \in Q.$$

⁴Recall that V^u is the cost-to-go associated with the control u and V_{θ} the NN parameterized by the weights θ .

⁵Since $V_{\theta}(y) \in C^2(\bar{Q})$, $\mathcal{H}'(y, \theta, \omega)$ is well a continuous function defined on $\bar{Q}$. From the compactness of $\bar{Q}$, we can thus always use Lemma 3.2 to approximate $\mathcal{H}'(y, \theta, \omega)$ for all $y \in \bar{Q}$ and hence for all $y \in Q$.

Similarly, from Lemma 6.1 and Assumption 6.2, there always exists $\omega^\#$ such that

$$\mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega^\#) = \min_{\omega} \mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega), \quad \forall y \in Q.$$

Therefore, since

$$\begin{aligned} \min_{\omega} \mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega) &= \min_{\omega} \mathbb{E}_{y \sim \mu_1(Q)} [\mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega)] \\ &\geq \mathbb{E}_{y \sim \mu_1(Q)} [\mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega^\#)] = \mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega^\#) \end{aligned}$$

and since by definition, $\min_{\omega} \mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega) \leq \mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega)$, $\forall \omega \in \mathbb{R}^p$, we have

$$\mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega^\#) = \min_{\omega} \mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega).$$

Then, since the global minima of $\mathcal{L}_2(\theta^{(k^\dagger+1)}, \omega)$ can be obtained in polynomial time, the modified Hamiltonian $\mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega)$ can be taken to global minimum for all $y \in Q$ by minimizing over ω . Therefore, we can also find $\omega^{(k^\dagger+1)}$ by applying successively (6.3.12) such that

$$\mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega^{(k^\dagger+1)}) \leq \mathcal{H}'(y, \theta^{(k^\dagger+1)}, \omega^{(k^\dagger)}) \leq 0, \quad \forall y \in Q. \quad (6.A.7)$$

Applying this repeatedly for subsequent steps, we have for all $k \geq k^\dagger$:

$$\mathcal{H}'(y, \theta^{(k)}, \omega^{(k)}) \leq \mathcal{H}'(y, \theta^{(k+1)}, \omega^{(k)}) \leq 0, \quad \forall y \in Q, \forall k \geq k^\dagger. \quad (6.A.8)$$

Then, since $V_\theta(y)$ is chosen $C^2(\bar{Q})$ for all $\theta \in \mathbb{R}^p$ (twice differentiable activation functions), we get by Dynkin's formula for all Markov controls $u := u(y)$,

$$\mathbb{E}_y^\mathbb{P} [V_{\theta^{(k)}}(Y_T^u)] = V_{\theta^{(k)}}(y) + \mathbb{E}_y^\mathbb{P} \left[\int_t^T (\mathcal{L}^u V_{\theta^{(k)}})(Y_r^u) dr \right],$$

which then leads to

$$V_{\theta^{(k)}}(y) = \mathbb{E}_y^\mathbb{P} \left[\phi(Y_T^u) - \int_t^T (\mathcal{L}^u V_{\theta^{(k)}})(Y_r^u) dr \right],$$

from Assumption 6.2. Similarly, for all Markov controls $u(y)$:

$$V_{\theta^{(k+1)}}(y) = \mathbb{E}_y^\mathbb{P} \left[\phi(Y_T^u) - \int_t^T (\mathcal{L}^u V_{\theta^{(k+1)}})(Y_r^u) dr \right].$$

Therefore,

$$V_{\theta^{(k+1)}}(y) - V_{\theta^{(k)}}(y) = \mathbb{E}_y^\mathbb{P} \left[\int_t^T ((\mathcal{L}^u V_{\theta^{(k)}})(Y_r^u) - (\mathcal{L}^u V_{\theta^{(k+1)}})(Y_r^u)) dr \right], \quad (6.A.9)$$

which is of course valid in particular for $u(y) = u_{\omega^{(k)}}(y)$ since it is admissible. However, from (6.A.8), we find for $k \geq k^\dagger$:

$$\begin{aligned} \mathcal{H}'(y, \theta^{(k)}, \omega^{(k)}) &\leq \mathcal{H}'(y, \theta^{(k+1)}, \omega^{(k)}), & \forall y \in Q, \\ \ell(y, u_{\omega^{(k)}}(y)) + (\mathcal{L}^{u_{\omega^{(k)}}} V_{\theta^{(k)}})(y) &\leq \ell(y, u_{\omega^{(k)}}(y)) + (\mathcal{L}^{u_{\omega^{(k)}}} V_{\theta^{(k+1)}})(y), & \forall y \in Q, \\ (\mathcal{L}^{u_{\omega^{(k)}}} V_{\theta^{(k)}})(y) &\leq (\mathcal{L}^{u_{\omega^{(k)}}} V_{\theta^{(k+1)}})(y), & \forall y \in Q, \end{aligned}$$

Then, since we have $V_{\theta}(y) \geq 0$, we find for $k \geq k^\dagger$

$$0 \leq V_{\theta^{(k+1)}}(y) \leq V_{\theta^{(k)}}(y), \quad \forall y \in \bar{Q}, \forall k \geq k^\dagger. \quad (6.A.10)$$

Therefore, $V_{\theta^{(k)}}(y)$ is pointwise convergent as k goes to ∞ and we can write $\lim_{k \rightarrow \infty} V_{\theta^{(k)}}(y) = V_{\theta^{(\infty)}}(y)$. From the compactness of $\bar{Q}$, we obtain uniform convergence in y by Dini's Theorem. This implies that

$$\lim_{k \rightarrow \infty} \sup_{y \in \bar{Q}} |V_{\theta^{(k)}}(y) - V_{\theta^{(k+1)}}(y)| = 0.$$

From (6.A.9), we find with $u(y) = u_{\omega^{(k)}}(y)$:

$$\begin{aligned} V_{\theta^{(k+1)}}(y) - V_{\theta^{(k)}}(y) &= \mathbb{E}_y \left[\int_t^T \left(\mathcal{H}'(Y_r^{u_{\omega^{(k)}}}, \theta^{(k)}, \omega^{(k)}) - \ell(Y_r^{u_{\omega^{(k)}}}, u_{\omega^{(k)}}(Y_r^{u_{\omega^{(k)}}})) \right. \right. \\ &\quad \left. \left. - \mathcal{H}'(Y_r^{u_{\omega^{(k)}}}, \theta^{(k+1)}, \omega^{(k)}) + \ell(Y_r^{u_{\omega^{(k)}}}, u_{\omega^{(k)}}(Y_r^{u_{\omega^{(k)}}})) \right) dr \right] \\ &= \mathbb{E}_y \left[\int_t^T \left(\mathcal{H}'(Y_r^{u_{\omega^{(k)}}}, \theta^{(k)}, \omega^{(k)}) - \mathcal{H}'(Y_r^{u_{\omega^{(k)}}}, \theta^{(k+1)}, \omega^{(k)}) \right) dr \right]. \end{aligned}$$

Hence, combined with inequality (6.A.8), this implies

$$\lim_{k \rightarrow \infty} \sup_{y \in \bar{Q}} \left| \mathcal{H}'(y, \theta^{(k+1)}, \omega^{(k)}) - \mathcal{H}'(y, \theta^{(k)}, \omega^{(k)}) \right| = 0. \quad (6.A.11)$$

Recall that we have $\min_{\theta} \mathcal{L}_1(\theta, \omega^{(k)}) = \min_{\theta} \mathbb{E}_{y \sim \mu_1(Q)} [\mathcal{H}'(y, \theta, \omega^{(k)})^2] = 0$, for all $k \geq k^\dagger$. Since $\theta^{(k+1)}$ is obtained by minimizing $\mathcal{L}_1(\theta^{(k)}, \omega^{(k)})$ using successively (6.3.11), then according to (6.A.11), we have

$$\lim_{k \rightarrow \infty} \mathcal{H}'(y, \theta^{(k)}, \omega^{(k)}) = \lim_{k \rightarrow \infty} \mathcal{H}'(y, \theta^{(k+1)}, \omega^{(k)}) = 0, \quad \forall y \in Q. \quad (6.A.12)$$

Therefore, $V_{\theta^{(\infty)}}(y)$ is solution of the Lyapunov equation (6.A.5) for a given policy $u_{\omega^{(\infty)}}(y)$ and hence,

$$V_{\theta^{(\infty)}}(y) = V^{u_{\omega^{(\infty)}}}(y), \quad \forall y \in Q. \quad (6.A.13)$$

where $V^{u_{\omega^{(\infty)}}}(y)$ is the cost-to-go function (6.A.1) associated with the Markov control $u_{\omega^{(\infty)}}(y)$. As shown in Abu-Khalaf and Lewis (2005), since $V_{\theta^{(k)}}(y)$ converges

uniformly to $V_{\theta^{(\infty)}}(y) = V^{u_{\omega^{(\infty)}}}(y)$ on $\bar{Q}$ and from the uniqueness of the state trajectories (see the Lipschitz continuity conditions on the dynamic (6.2.2)), the sequence of controls $u_{\omega^{(k)}}(y)$ converges uniformly to $u_{\omega^{(\infty)}}(y)$ on Q when $k \rightarrow \infty$. Since $\omega^{(k+1)}$ is obtained by minimizing $\mathcal{L}_2(\theta^{(k+1)}, \omega)$ using (6.3.12), it is obvious from the dominated convergence theorem with (6.A.8) and from (6.A.12) that

$$\lim_{k \rightarrow \infty} \mathcal{L}_2(\theta^{(k+1)}, \omega^{(k+1)}) = \lim_{k \rightarrow \infty} \mathcal{L}_2(\theta^{(k+1)}, \omega^{(k)}) = 0, \quad (6.A.14)$$

which implies that

$$\lim_{k \rightarrow \infty} \min_{\omega} \mathcal{H}'(y, \theta^{(k)}, \omega) = 0, \quad \forall y \in Q. \quad (6.A.15)$$

From (6.A.12), (6.A.15) and Theorem 6.2, one has $\lim_{k \rightarrow \infty} V_{\theta^{(k)}}(y) = V(y)$ for all $y \in \bar{Q}$ and $\lim_{k \rightarrow \infty} u_{\omega^{(k)}}(y) = u^*(y)$ for all $y \in Q$. Therefore, we conclude that $V_{\theta^{(k)}}(y)$ goes to the value function $V(y)$ uniformly on $\bar{Q}$ and $u_{\omega^{(k)}}(y)$ to the optimal control $u^*(y)$ uniformly on Q as $k \rightarrow \infty$, which means the global optimal solution is achieved. $\square$

Conclusion and extensions

Liquidity is a complex – yet important – concept at the heart of many financial crises and disasters. This thesis therefore attempts to provide a better understanding of liquidity using various tools ranging from rough volatility, subdiffusions, fractional Hawkes, optimal control and deep learning. As the definition of liquidity depends on the temporal granularity at which it is analyzed (high-frequency, low-frequency and intra-day time scales), this dissertation addresses liquidity modeling in three distinct parts, each corresponding to one of these three temporal perspectives along with their connections.

The first part of the thesis aims at studying rough volatility models (and particularly the rough Heston model) as they arise from the modeling of market liquidity at high-frequency and since they are excellent candidates for reproducing important stylized facts of both the past volatility time series and the implied volatility surface. In Chapter 1, we explore more closely two tractable implementations of the RFSV model, the rBergomi and rough Heston models. We show how to expand these two rough volatility models at larger time scales, we analyze their implications for the pricing of long-term life insurance claims and we explain why they provide a more accurate fair value of such long-term contracts. Compared to standard stochastic volatility models, we conclude in this chapter that the rough Heston model provides efficiently a more accurate fair value of long-term life insurance contracts while being highly consistent with historical and risk-neutral data, thereby contributing to bridging the gap between high-frequency liquidity modeling and long-term asset/insurance pricing. Chapter 2 then studies this rough Heston model (and more generally affine Volterra processes) in the context of portfolio insurance strategies. We indeed propose an approach based on the characteristic function of the log-cushion, which appears to be more consistent and particularly efficient in the case of affine Volterra processes compared with existing simulation

techniques. Using such approach, we describe several properties of CPPI which naturally result from the log-cushion's characteristic function under affine Volterra processes. This allows to consider more realistic dynamics for the underlying risky asset in the context of CPPI so as to build portfolio strategies that are more consistent with financial data and better suited to handle the liquidity risk associated with CPPI.

Our focus then shifts in the second part of the thesis to directly modeling the illiquidity process associated with low-frequency observations of asset prices using fractional processes. Chapter 3 indeed proposes a new subdiffusive model of thinly-traded asset prices reproducing the main characteristics exhibited in such illiquid markets. This is done by considering a stochastic volatility jump model, time-changed by an inverse subordinator driving the (il)liquidity process of the related asset. We then derive the forward fractional PDEs governing the probability density function of the introduced model and we prove that it leads to an arbitrage-free and incomplete market under a suitable change of measure. By proposing a new procedure for estimating the model parameters and using a series expansion for solving numerically the obtained fractional PDE, we are able to price various European-type derivatives on illiquid assets and to depart from the common Markovian valuation setup. Chapter 4 then proposes a new paradigm for modeling illiquidity via a novel fractional Hawkes process in which the intensity process is ruled by a modified Mittag-Leffler excitation function. By considering a mean-reverting jump diffusion for the (log-)Amihud measure with jumps following this modified fractional Hawkes process, we then manage to effectively reproduce the observed peaks of illiquidity in financial markets while introducing long-memory and tractability in the model. By deriving the characteristic function satisfied by the Amihud measure in this model, we provide in the last part of the chapter an efficient way to perform risk management on illiquidity. These two Chapters 3 and 4 hence provide new tools based on fractional processes for a better management of the illiquidity risk in financial markets at the low-frequency level.

The third part of the thesis finally focus on optimal liquidation problems in finance as they are built upon price impact models and thus consider liquidity modeling from an intermediate (intraday) time scale. Chapter 5 first proposes a new framework of optimal liquidation strategies for a trader seeking to liquidate his large inventory based on a jump-dependent price impact model with propagator. This new jump-dependent price impact model best reproduces the empirical direct and indirect effects of market orders on the transaction price by introducing a stochastic liquidity factor in the form of a trade-driven intensity process. We conclude with Chapter 5 that the introduced jump-dependent price impact model leads to an alternative, realistic and flexible description of the order book's dynamics and its liquidity. At last, Chapter 6 introduces in a more general way the GPI-PINN algorithm, a novel numerical scheme for solving continuous-time stochastic optimal

control problems in high dimensions when the optimal control does not admit an explicit solution. Combining Physics-Informed Neural Networks with an Actor-Critic structure built upon the Generalized Policy Iteration technique, this successive deep learning algorithm employs two separate neural networks to approximate both the value function and the multidimensional optimal control, leading to a global approximation of the optimal solution across all time and space. Chapter 6 then proves the theoretical optimality and convergence guarantees of the scheme. Moreover, its accuracy and efficacy are shown empirically based on two numerical examples, and in particular for optimal liquidation strategies with numerous co-integrated assets and without explicit solutions for the optimal control.

Several research subjects around liquidity, fractional processes and optimal control have been addressed throughout this dissertation, although many avenues have not been fully explored. At the high-frequency level, high-frequency price models have only been investigated in this thesis through their connections with long-term asset pricing in Chapter 1 and with price impact models in Chapter 5. Fractional processes could be used instead to directly reproduce the empirical properties of high-frequency markets and their liquidity. In particular, we should try to fit the fractional Hawkes process derived in Chapter 4 to high-frequency data and see whether it manages to better replicate their statistical properties, compared with existing high-frequency price models. Similarly, the subdiffusive SVJ model in Chapter 3 could also be used to reproduce the sample paths of liquid assets observed at very high frequency. Indeed, freeze periods and liquidity shortages such as in Figure 3.1.1 also occur in most stock prices at very short time scales, even for liquid assets, due to the discrete nature of time trade arrivals and price variation. The Markovian representation of subdiffusive assets obtained in Chapter 3 (conditionally on the estimation of S_s, U_{S_s}) should then allow us to perform optimal control and hence derive optimal liquidation strategies based on such subdiffusive dynamics. Moreover, other than asset liquidity, using these fractional processes for modeling interest rates and credit risk is a way to explore. Finally, the algorithm developed for high-dimensional implicit optimal control in Chapter 6 is quite general and could be applied to other fields in liquidity modeling than optimal liquidation, such as market making, pairs trading or asset and liability management.

References

- Frédéric Abergel and Aymen Jedidi. A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance*, 16(05):1350025, 2013.
- Eduardo Abi Jaber. Lifting the Heston model. *Quantitative Finance*, 19(12):1995–2013, 2019.
- Eduardo Abi Jaber. Weak existence and uniqueness for affine stochastic Volterra equations with L1-kernels. *Bernoulli*, 27(3):1583–1615, 2021.
- Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019.
- Eduardo Abi Jaber and Eyal Neuman. Optimal liquidation with signals: the general propagator case. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4264823>, 2022.
- Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *Annals of Applied Probability*, 29(5):3155–3200, 2019.
- Murad Abu-Khalaf and Frank L Lewis. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 41(5):779–791, 2005.
- Julia Ackermann, Thomas Kruse, and Ludger Overbeck. Inhomogeneous affine Volterra processes. *arXiv preprint arXiv:2012.10966*, 2020.
- Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.

- Mesias Alfeus, Christina S Nikitopoulos, and Ludger Overbeck. Implied roughness in the term structure of oil market volatility. *Quantitative Finance*, pages 1–17, 2024.
- Aurélien Alfonsi and Pierre Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics*, 20(1):183–218, 2016.
- Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.
- Zeyuan Allen-Zhu, Yuezhi Li, and Zhao Song. A convergence theory for deep learning via over-parameterization. In *International conference on machine learning*, pages 242–252. PMLR, 2019.
- Robert Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18, 2003.
- Robert Almgren and Neil Chriss. Value under liquidation. *Risk*, 12(12):61–63, 1999.
- Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3(2):5–40, 2001.
- Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.
- Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, 2002.
- Yakov Amihud. The pricing of illiquidity as a characteristic and as risk. *Multinational Finance Journal*, 19(3):149–168, 2015.
- Leif BG Andersen and Vladimir V Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.
- Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Heiko Ebens. The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76, 2001.
- Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.
- David Applebaum. *Lévy processes and Stochastic Calculus*. Cambridge University Press, 2009.
- Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

- Yacine Aït-Sahalia and Thomas Robert Hurd. Portfolio choice in markets with contagion. *Journal of Financial Econometrics*, 14(1):1–28, 01 2016.
- Achref Bachouch, Côme Huré, Nicolas Langrené, and Huyen Pham. Deep neural networks algorithms for stochastic control problems on finite horizon: numerical applications. *Methodology and Computing in Applied Probability*, 24(1):143–178, 2022.
- Anna Rita Bacinello, Enrico Biffis, and Pietro Millossovich. Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance*, 10(9):1077–1090, 2010.
- Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.
- Emmanuel Bacry, Khalil Dayri, and Jean-François 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.
- Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.
- Yuxing Bai, Temuer Chaolu, and Sudao Bilige. The application of improved physics-informed neural network (IPINN) method in finance. *Nonlinear Dynamics*, 107(4):3655–3667, 2022.
- Sven Balder, Michael Brandl, and Antje Mahayni. Effectiveness of CPPI strategies under discrete-time trading. *Journal of Economic Dynamics and Control*, 33(1):204–220, 2009.
- Harry Bateman. *Higher Transcendental Functions*, volume 1. McGraw-Hill Book Company, New-York, 1953.
- David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *The Review of Financial Studies*, 9(1):69–107, 1996.
- Nicole Bäuerle and Sascha Desmettre. Portfolio optimization in fractional and rough Heston models. *SIAM Journal on Financial Mathematics*, 11(1):240–273, 2020.
- Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- Christian Bayer, Peter K Friz, Archil Gulisashvili, Blanka Horvath, and Benjamin Stemper. Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, 19(5):779–798, 2019, a.

- Christian Bayer, Blanka Horvath, Aitor Muguruza, Benjamin Stemper, and Mehdi Tomas. On deep calibration of (rough) stochastic volatility models. *arXiv preprint arXiv:1908.08806*, 2019, b.
- Erhan Bayraktar and Michael Ludkovski. Liquidation in limit order books with controlled intensity. *Mathematical Finance*, 24(4):627–650, 2014.
- Christian Beck, Sebastian Becker, Philipp Grohs, Nor Jaafari, and Arnulf Jentzen. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv preprint arXiv:1806.00421*, 1(1), 2018.
- Christian Beck, Weinan E, and Arnulf Jentzen. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *Journal of Nonlinear Science*, 29:1563–1619, 2019.
- Christian Beck, Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Ariel Neufeld. Deep splitting method for parabolic PDEs. *SIAM Journal on Scientific Computing*, 43(5):A3135–A3154, 2021.
- Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. *arXiv preprint arXiv:1610.00332*, 2016.
- Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.
- Fred Espen Benth, Asma Khedher, and Michèle Vanmaele. Pricing of commodity derivatives on processes with memory. *Risks*, 8(1):8, 2020.
- Jens Berg and Kaj Nyström. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing*, 317:28–41, 2018.
- Lorenzo Bergomi. *Stochastic Volatility Modeling*. CRC press, 2015.
- Lorenzo Bergomi and Julien Guyon. Stochastic volatility’s orderly smiles. *Risk*, 25(5):60, 2012.
- Julius Berner, Markus Dablander, and Philipp Grohs. Numerically solving parametric families of high-dimensional Kolmogorov partial differential equations via deep learning. *Advances in Neural Information Processing Systems*, 33:16615–16627, 2020.
- Jean Bertoin. *Subordinators: Examples and Applications*, pages 1–91. Springer Berlin Heidelberg, 1999.

- Philippe Bertrand and Jean-Luc Prigent. Portfolio insurance strategies: OBPI versus CPPI. In *Conférence Internationale en Finance*, 2001.
- Philippe Bertrand and Jean-Luc Prigent. Portfolio insurance: The extreme value approach applied to the CPPI method. *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*, 465, 2016.
- Philippe Bertrand and Jean-Luc Prigent. On the optimality of path-dependent structured funds: The cost of standardization. *European Journal of Operational Research*, 277(1):333–350, 2019.
- Dimitri Bertsekas. *A Course in Reinforcement Learning*. Athena Scientific, 2023.
- Dimitris Bertsimas and Andrew W Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.
- Necdet Bildik, Ali Konuralp, Funda Orakçı Bek, and Semih Küçükarslan. Solution of different type of the partial differential equation by differential transform method and Adomian’s decomposition method. *Applied Mathematics and Computation*, 172(1):551–567, 2006.
- Fischer Black and Robert Jones. Simplifying portfolio insurance. *Journal of portfolio management*, 14(1):48, 1987.
- Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.
- Jan Blechschmidt and Oliver G Ernst. Three ways to solve partial differential equations with neural networks—a review. *GAMM-Mitteilungen*, 44(2):e202100006, 2021.
- Tim Bollerslev, Michael Gibson, and Hao Zhou. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of econometrics*, 160(1):235–245, 2011.
- Alessandro Bondi, Sergio Pulido, and Simone Scotti. The rough Hawkes Heston stochastic volatility model. *arXiv preprint arXiv:2210.12393*, 2022.
- Alessandro Bondi, Giulia Livieri, and Sergio Pulido. Affine Volterra processes with jumps. *Stochastic Processes and their Applications*, 168:104264, 2024.
- Bruno Bouchard. Introduction to stochastic control of mixed diffusion processes, viscosity solutions and applications in finance and insurance. *Lecture Notes Preprint*, 2007.
- Bruno Bouchard and Nizar Touzi. Weak dynamic programming principle for viscosity solutions. *SIAM Journal on Control and Optimization*, 49(3):948–962, 2011.

- Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of random price changes. *Quantitative Finance*, 4(2):176, 2003.
- Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, Quotes and Prices: financial markets under the microscope*. Cambridge University Press, 2018.
- Nicole Branger, Antje Mahayni, and Judith C Schneider. On the optimal design of insurance contracts with guarantees. *Insurance: Mathematics and Economics*, 46(3):485–492, 2010.
- Michael Brennan, Sahn-Wook Huh, and Avanidhar Subrahmanyam. An analysis of the Amihud illiquidity premium. *The Review of Asset Pricing Studies*, 3(1):133–176, 2013.
- Hans Buehler. Consistent variance curve models. *Finance and Stochastics*, 10(2):178–203, 2006.
- Marinella Cadoni, Roberta Melis, and Alessandro Trudda. Pension funds rules: Paradoxes in risk control. *Finance Research Letters*, 22:20–29, 2017.
- Andrew JG Cairns, David Blake, and Kevin Dowd. Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, 2008(2-3):79–113, 2008.
- Charles Cao, Yong Chen, Bing Liang, and Andrew W Lo. Can hedge funds time market liquidity? *Journal of Financial Economics*, 109(2):493–516, 2013.
- Guglielmo Maria Caporale, Luis Gil-Alana, and Alex Plastun. Is market fear persistent? A long-memory analysis. *Finance Research Letters*, 27:140–147, 2018.
- Peter Carr and Dilip Madan. Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.
- Peter Carr and Liuren Wu. The finite moment log stable process and option pricing. *The journal of finance*, 58(2):753–777, 2003.
- Álvaro Cartea and Diego del Castillo-Negrete. Fractional diffusion models of option prices in markets with jumps. *Physica A: Statistical Mechanics and its Applications*, 374(2):749–763, 2007.
- Álvaro Cartea and Sam Howison. Option pricing with Lévy-stable processes generated by Lévy-stable integrated variance. *Quantitative Finance*, 9(4):397–409, 2009.

- Álvaro Cartea and Sebastian Jaimungal. Incorporating order-flow into optimal execution. *Mathematics and Financial Economics*, 10:339–364, 2016.
- Álvaro Cartea and Thilo Meyer-Brandis. How duration between trades of underlying securities affects option prices. *Review of Finance*, 14(4):749–785, 2010.
- Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014.
- Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and High-Frequency Trading*. Cambridge University Press, 2015.
- Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. Algorithmic trading with model uncertainty. *SIAM Journal on Financial Mathematics*, 8(1):635–671, 2017.
- Alvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. Enhancing trading strategies with order book signals. *Applied Mathematical Finance*, 25(1):1–35, 2018.
- Álvaro Cartea, Luhui Gan, and Sebastian Jaimungal. Trading co-integrated assets with price impact. *Mathematical Finance*, 29(2):542–567, 2019.
- Thomas Cayé and Johannes Muhle-Karbe. Liquidation with self-exciting price impact. *Mathematics and Financial Economics*, 10:15–28, 2016.
- Quentin Chan-Wai-Nam, Joseph Mikael, and Xavier Warin. Machine learning for semi linear PDEs. *Journal of scientific computing*, 79(3):1667–1712, 2019.
- Jing Chen, AG Hawkes, and E Scalas. A fractional Hawkes process. In *Nonlocal and Fractional Operators*, pages 121–131. Springer, New-York, 2021.
- Xia Chen, Wenbo V Li, Jan Rosiński, and Qi-Man Shao. Large deviations for local times and intersection local times of fractional Brownian motions and Riemann–Liouville processes. *The Annals of Probability*, 39(2):729–778, 2011.
- Patrick Cheridito. Arbitrage in fractional Brownian motion models. *Finance and Stochastics*, 7(4):533–553, 2003.
- Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of probability*, 8, 2003.
- Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.
- Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.

- Rama Cont and Adrien De Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.
- Rama Cont and Peter Tankov. Constant proportion portfolio insurance in the presence of jumps in asset prices. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(3):379–401, 2009.
- Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014.
- Laure Coutin and Monique Pontier. Approximation of the fractional Brownian sheet via Ornstein-Uhlenbeck sheet. *ESAIM: Probability and Statistics*, 11:115–146, 2007.
- Yiran Cui, Sebastian del Bano Rollin, and Guido Germano. Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research*, 263(2):625–638, 2017.
- Gianbiagio Curato, Jim Gatheral, and Fabrizio Lillo. Optimal execution with non-linear transient market impact. *Quantitative Finance*, 17(1):41–54, 2017.
- José Da Fonseca and Yannick Malevergne. A simple microstructure model based on the cox-BESQ process with application to optimal execution policy. *Journal of Economic Dynamics and Control*, 128:104137, 2021.
- José Da Fonseca and Riadh Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34(6):548–579, 2014.
- Tim De Ryck and Siddhartha Mishra. Error analysis for Physics-Informed Neural Networks (PINNs) approximating Kolmogorov PDEs. *Advances in Computational Mathematics*, 48(6):79, 2022.
- Kristian Debrabant and Espen R Jakobsen. Semi-Lagrangian schemes for linear and fully non-linear diffusion equations. *Mathematics of Computation*, 82(283):1433–1462, 2013.
- Jan Dhaene, Steven Vanduffel, Marc J Goovaerts, Rob Kaas, Qihe Tang, and David Vyncke. Risk measures and comonotonicity: a review. *Stochastic Models*, 22(4):573–606, 2006.
- Kai Diethelm, Neville J Ford, and Alan D Freed. A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dynamics*, 29(1):3–22, 2002.
- Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

- Ryan Donnelly. Optimal execution: A review. *Applied Mathematical Finance*, 29(3):181–212, 2022.
- Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.
- Adrian A Drăgulescu and Victor M Yakovenko. Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative Finance*, 2(6):443–453, 2002.
- Simon Du, Jason Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai. Gradient descent finds global minima of deep neural networks. In *International conference on machine learning*, pages 1675–1685. PMLR, 2019.
- Jingliang Duan, Jie Li, Qiang Ge, Shengbo Eben Li, Monimoy Bujarbaruah, Fei Ma, and Dezhao Zhang. Relaxed Actor-Critic with convergence guarantees for continuous-time optimal control of nonlinear systems. *IEEE Transactions on Intelligent Vehicles*, 8(5):3299–3311, 2023.
- Alfonso Dufour and Robert F Engle. Time and the price impact of a trade. *The Journal of Finance*, 55(6):2467–2498, 2000.
- Jean-Loup Dupret and Jérôme Barbarin. Rough stochastic volatility modeling. *Faculté des sciences, Université catholique de Louvain*, 2020.
- Jean-Loup Dupret and Donatien Hainaut. Portfolio insurance under rough volatility and Volterra processes. *International Journal of Theoretical and Applied Finance*, 24(06n07):2150036, 2021.
- Jean-Loup Dupret and Donatien Hainaut. A fractional Hawkes process for illiquidity modeling. Technical report, LIDAM Discussion Paper ISBA, 2023a.
- Jean-Loup Dupret and Donatien Hainaut. Optimal liquidation under indirect price impact with propagator. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4625578>, 2023b.
- Jean-Loup Dupret and Donatien Hainaut. A subdiffusive stochastic volatility jump model. *Quantitative Finance*, 23(6):979–1002, 2023c.
- Jean-Loup Dupret, Jérôme Barbarin, and Donatien Hainaut. Impact of rough stochastic volatility models on long-term life insurance pricing. *European Actuarial Journal*, 13(1):235–275, 2023.
- David Easley and Maureen O’hara. Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1):69–90, 1987.

- Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.
- Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.
- Omar El Euch, Masaaki Fukasawa, and Mathieu Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22: 241–280, 2018.
- Omar El Euch, Jim Gatheral, and Mathieu Rosenbaum. Roughening Heston. *Risk*, pages 84–89, 2019.
- Paul Embrechts, Thomas Liniger, and Lu Lin. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378, 2011.
- Eymen Errais, Kay Giesecke, and Lisa R Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.
- Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.
- Martin Forde and Hongzhong Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1):114–145, 2017.
- Peter A Forsyth, J Shannon Kennedy, Shu Tong Tse, and Heath Windcliff. Optimal trade execution: a mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 36(12):1971–1991, 2012.
- Jean-Pierre Fouque, Sebastian Jaimungal, and Yuri F Saporito. Optimal trading with signals and stochastic price impact. *SIAM Journal on Financial Mathematics*, 13(3):944–968, 2022.
- Peter K Friz and Martin Keller Ressel. Moment explosions. In *Encyclopedia of Quantitative Finance*, pages 1247–1253. John Wiley & Sons, 2010. ISBN 9780470061602.
- Antje Fruth, Torsten Schöneborn, and Mikhail Urusov. Optimal trade execution and price manipulation in order books with time-varying liquidity. *Mathematical Finance*, 24(4):651–695, 2014.
- Antje Fruth, Torsten Schöneborn, and Mikhail Urusov. Optimal trade execution in order books with stochastic liquidity. *Mathematical Finance*, 29(2):507–541, 2019.

- Masaaki Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.
- Ahcène Gareche, Gaétan Disdier, Julianus Kockelkoren, and Jean-Philippe Bouchaud. Fokker-planck description for the queue dynamics of large tick stocks. *Physical Review E*, 88(3):032809, 2013.
- Jim Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, 2010.
- Jim Gatheral. *The volatility surface: a practitioner’s guide*, volume 357. John Wiley & Sons, 2011.
- Jim Gatheral and Antoine Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.
- Jim Gatheral and Martin Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23(3):501–533, 2019.
- Jim Gatheral and Radoš Radoičić. Rational approximation of the rough Heston solution. *International Journal of Theoretical and Applied Finance*, 22(03):1950010, 2019.
- Jim Gatheral and Alexander Schied. *Dynamical models of market impact and algorithms for order execution*, pages 579–599. Cambridge University Press, 2013.
- Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative finance*, 18(6):933–949, 2018.
- Stefan Gerhold, Christoph Gerstenecker, and Arpad Pinter. Moment explosions in the rough Heston model. *Decisions in Economics and Finance*, 42(2):575–608, 2019.
- Kathrin Glau and Linus Wunderlich. The deep parametric PDE method and applications to option pricing. *Applied Mathematics and Computation*, 432:127355, 2022.
- Selim Gökay, Alexandre F. Roch, and H. Mete Soner. *Liquidity Models in Continuous and Discrete Time*, pages 333–365. Springer Berlin Heidelberg, 2011.
- Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F-radar and signal processing*, volume 140, pages 107–113. IET, 1993.
- Ruslan Y Goyenko, Craig W Holden, and Charles A Trzcinka. Do liquidity measures measure liquidity? *Journal of financial Economics*, 92(2):153–181, 2009.

- Paulwin Graewe and Ulrich Horst. Optimal trade execution with instantaneous price impact and stochastic resilience. *SIAM Journal on Control and Optimization*, 55(6):3707–3725, 2017.
- Constantin Greif. Numerical methods for Hamilton-Jacobi-Bellman equations. 2017. URL <https://api.semanticscholar.org/CorpusID:125438964>.
- Benedict Gross. Note on the inversion of the Laplace transform. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(317):543–544, 1950.
- Sanford J Grossman and Jean-Luc Vila. Optimal dynamic trading with leverage constraints. *Journal of Financial and Quantitative Analysis*, pages 151–168, 1992.
- Hui Gu, Jin-Rong Liang, and Yun-Xiu Zhang. Time-changed geometric fractional Brownian motion and option pricing with transaction costs. *Physica A: Statistical Mechanics and its Applications*, 391(15):3971–3977, 2012.
- Olivier Guéant. *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, volume 33. CRC Press, 2016.
- Olivier Guéant and Charles-Albert Lehalle. General intensity shapes in optimal liquidation. *Mathematical Finance*, 25(3):457–495, 2015.
- Zhidong Guo and Hongjun Yuan. Pricing European option under the time-changed mixed Brownian-fractional Brownian model. *Physica A: Statistical Mechanics and its Applications*, 406:73–79, 2014.
- Christian Hafner, Oliver B Linton, and Linqi Wang. Dynamic Autoregressive Liquidity (DARLiQ). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4042069>, 2022.
- Donatien Hainaut. Risk management of CPPI funds in switching regime markets. *Bankers, Markets and Investors*. August, 2011.
- Donatien Hainaut. A model for interest rates with clustering effects. *Quantitative Finance*, 16(8):1203–1218, 2016.
- Donatien Hainaut. Fractional Hawkes processes. *Physica A: Statistical Mechanics and its Applications*, 549:124330, 2020.
- Donatien Hainaut. A fractional multi-states model for insurance. *Insurance: Mathematics and Economics*, 98:120–132, 2021.
- Donatien Hainaut. *Continuous time processes for finance. Switching, self-exciting, fractional and other recent dynamics*. Bocconi & Springer Series, London, 2022.

- Donatien Hainaut. Valuation of guaranteed minimum accumulation benefits (GMAB) with physics inspired neural networks. *LIDAM Discussion Paper ISBA*, 2023(29), 2023.
- Donatien Hainaut and Pierre Devolder. Mortality modelling with Lévy processes. *Insurance: Mathematics and Economics*, 42(1):409–418, 2008.
- Donatien Hainaut and Stephane Goutte. A switching microstructure model for stock prices. *Mathematics and Financial Economics*, 13:459–490, 2019.
- Donatien Hainaut and Nikolai Leonenko. Option pricing in illiquid markets: A fractional jump–diffusion approach. *Journal of Computational and Applied Mathematics*, 381:112995, 2021.
- Donatien Hainaut and Franck Moraux. Hedging of options in the presence of jump clustering. *Journal of Computational Finance, Forthcoming*, 2018.
- Jiequn Han and Weinan E. Deep learning approximation for stochastic control problems. *arXiv preprint arXiv:1611.07422*, 2016.
- Jiequn Han, Arnulf Jentzen, and Weinan E. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.
- Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.
- Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.
- Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- Alan G Hawkes. Hawkes jump-diffusions and finance: a brief history and review. *The European Journal of Finance*, 28(7):627–641, 2022.
- Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.
- Isidore Isaac Hirschman and David V Widder. *The convolution transform*. Courier Corporation, Chelmsford, Massachusetts, US, 2012.
- Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

- Ronald A Howard. *Dynamic programming and Markov processes*. John Wiley, 1960.
- Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.
- Gur Huberman and Werner Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.
- Côme Huré, Huyên Pham, Achref Bachouch, and Nicolas Langrené. Deep neural networks algorithms for stochastic control problems on finite horizon: convergence analysis. *SIAM Journal on Numerical Analysis*, 59(1):525–557, 2021.
- Simon R Hurst, Eckhard Platen, and Svetlozar Todorov Rachev. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29(10-12):105–119, 1999.
- Martin Hutzenthaler, Arnulf Jentzen, Thomas Kruse, and Tuan Anh Nguyen. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *SN partial differential equations and applications*, 1:1–34, 2020.
- Mariya Ishteva, Lyubomir Boyadjiev, and Rudolf Scherer. On the Caputo operator of fractional calculus and C-Laguerre functions. *Mathematical Sciences Research Journal*, 9(6):161, 2005.
- Saul D Jacka and Aleksandar Mijatović. On the policy improvement algorithm in continuous time. *Stochastics*, 89(1):348–359, 2017.
- Kris Jacobs and Yuguo Liu. Estimation and filtering with big option data. *Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3300564>*, 2018.
- Jean Jacod. *Calcul stochastique et problèmes de martingales*, volume 714. Springer, 1979.
- Antoine Jacquier and Lorenzo Torricelli. Anomalous diffusions in option prices: connecting trade duration and the volatility term structure. *SIAM Journal on Financial Mathematics*, 11(4):1137–1167, 2020.
- Thibault Jaisson and Mathieu Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 26(5):2860–2882, 2016.
- J Janczura and A Wyłomańska. Subdynamics of financial data from fractional Fokker-Planck equation. *Acta Physica Polonica. Series B*, 40(5):1341–1351, 2009.

- Max Jensen and Iain Smears. On the convergence of finite element methods for Hamilton–Jacobi–Bellman equations. *SIAM Journal on Numerical Analysis*, 51(1):137–162, 2013.
- Arnulf Jentzen, Benno Kuckuck, Ariel Neufeld, and Philippe von Wurstemberger. Strong error analysis for stochastic gradient descent optimization algorithms. *IMA Journal of Numerical Analysis*, 41(1):455–492, 2021.
- Shaolin Ji, Shige Peng, Ying Peng, and Xichuan Zhang. Solving stochastic optimal control problem via stochastic maximum principle with deep learning method. *Journal of Scientific Computing*, 93(1):30, 2022.
- Deqing Jiang, Justin Sirignano, and Samuel N Cohen. Global convergence of Deep Galerkin and PINNs methods for solving partial differential equations. *arXiv preprint arXiv:2305.06000*, 2023.
- Michael Johannes and Nicholas Polson. MCMC methods for continuous-time financial econometrics. In *Handbook of Financial Econometrics: Applications*, pages 1–72. Elsevier, 2010.
- Michael S Johannes, Nicholas G Polson, and Jonathan R Stroud. Optimal filtering of jump diffusions: Extracting latent states from asset prices. *The Review of Financial Studies*, 22(7):2759–2799, 2009.
- Agnieszka Jurlewicz, Peter Kern, Mark M Meerschaert, and Hans-Peter Scheffler. Fractional governing equations for coupled random walks. *Computers & Mathematics with Applications*, 64(10):3021–3036, 2012.
- Ioannis Karatzas and Steven E Shreve. Brownian motion. In *Brownian Motion and Stochastic Calculus*, pages 47–127. Springer, 1998.
- Chris Kenyon and Andrew Green. Dirac Processes and Default Risk. *arXiv preprint arXiv:1504.04581*, 2015.
- Geoffrey Kingston. Theoretical foundations of constant-proportion portfolio insurance. *Economics Letters*, 29(4):345–347, 1989.
- Kei Kobayashi. Stochastic calculus for a time-changed semimartingale and the associated stochastic differential equations. *Journal of Theoretical Probability*, 24(3):789–820, 2011.
- Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.
- Andreas E Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Springer Science & Business Media, New-York, 2006.

- Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017.
- Gwo Dong Lin. On the Mittag–Leffler distributions. *Journal of Statistical Planning and Inference*, 74(1):1–9, 1998.
- Ignacio N Lobato and Carlos Velasco. Long memory in stock-market trading volume. *Journal of Business & Economic Statistics*, 18(4):410–427, 2000.
- Roger Lord, Remmert Koekkoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010.
- Xiaoxia Lou and Tao Shu. Price impact or trading volume: Why is the Amihud (2002) measure priced? *The Review of Financial Studies*, 30(12):4481–4520, 2017.
- Bing Lu. Optimal selling of an asset with jumps under incomplete information. *Applied Mathematical Finance*, 20(6):599–610, 2013.
- Marcin Magdziarz. Black-Scholes formula in subdiffusive regime. *Journal of Statistical Physics*, 136(3):553–564, 2009a.
- Marcin Magdziarz. Stochastic representation of subdiffusion processes with time-dependent drift. *Stochastic Processes and their Applications*, 119(10):3238–3252, 2009b.
- Francesco Mainardi. Why the Mittag-Leffler function can be considered the queen function of the fractional calculus? *Entropy*, 22(12):1359, 2020.
- Francesco Mainardi, Marco Raberto, Rudolf Gorenflo, and Enrico Scalas. Fractional calculus and continuous-time finance II: the waiting-time distribution. *Physica A: Statistical Mechanics and its Applications*, 287(3-4):468–481, 2000.
- Benoit B Mandelbrot and John W Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
- Ryan McCrickerd and Mikko S Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11):1877–1886, 2018.
- Mark M Meerschaert and Enrico Scalas. Coupled continuous time random walks in finance. *Physica A: Statistical Mechanics and its Applications*, 370(1):114–118, 2006.
- Mark M Meerschaert and Hans-Peter Scheffler. Limit theorems for continuous-time random walks with infinite mean waiting times. *Journal of applied probability*, 41(3):623–638, 2004.

- Mark M Meerschaert and Hans-Peter Scheffler. Triangular array limits for continuous time random walks. *Stochastic processes and their applications*, 118(9):1606–1633, 2008.
- Mark M Meerschaert and Peter Straka. Semi-Markov approach to continuous time random walk limit processes. *The Annals of Probability*, 42(4):1699–1723, 2014.
- Ralf Metzler and Joseph Klafter. The random walk’s guide to anomalous diffusion: a fractional dynamics approach. *Physics reports*, 339(1):1–77, 2000.
- Ralf Metzler and Joseph Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *Journal of Physics A: Mathematical and General*, 37(31):R161, 2004.
- Shaher Momani and Nabil Shawagfeh. Decomposition method for solving fractional Riccati differential equations. *Applied Mathematics and Computation*, 182(2):1083–1092, 2006.
- Saviz Mowlavi and Saleh Nabi. Optimal control of PDEs using physics-informed neural networks. *Journal of Computational Physics*, 473:111731, 2023.
- Thibaut Moyaert and Mikael Petitjean. The performance of popular stochastic volatility option pricing models during the subprime crisis. *Applied Financial Economics*, 21(14):1059–1068, 2011.
- Milan Mrázek, Jan Pospíšil, and Tomáš Sobotka. On calibration of stochastic and fractional stochastic volatility models. *European Journal of Operational Research*, 254(3):1036–1046, 2016.
- Marek Musiela and Marek Rutkowski. *Martingale Methods in financial Modelling*, volume 36. Springer Science & Business Media, 2006.
- Tenavi Nakamura-Zimmerer, Qi Gong, and Wei Kang. Adaptive deep learning for high-dimensional Hamilton–Jacobi–Bellman equations. *SIAM Journal on Scientific Computing*, 43(2):A1221–A1247, 2021.
- Syoiti Ninomiya and Nicolas Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, 15(2):107–121, 2008.
- Charles G Njike Leunga. Long memory self-exciting jump diffusion for asset prices modeling. Technical report, LIDAM Discussion Paper ISBA ; 2022/03, 2022.
- Charles Guy Njike Leunga and Donatien Hainaut. Affine Heston model style with self-exciting jumps and long memory. *Annals of Finance*, pages 1–43, 2024.
- Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

- Zaid M Odibat and Nabil T Shawagfeh. Generalized Taylor's formula. *Applied Mathematics and Computation*, 186(1):286–293, 2007.
- Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- Darren Pain and Jonathan Rand. Recent developments in portfolio insurance. *Recent Developments in Portfolio Insurance*, 2008.
- Louis Paulot and Xavier Lacroze. Efficient pricing of CPPI using Markov operators. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.1326297>, 2009.
- Andre F Perold and William F Sharpe. Dynamic strategies for asset allocation. *Financial Analysts Journal*, 44(1):16–27, 1988.
- Helfried Peyrl, Florian Herzog, and Hans P Geering. Numerical solution of the Hamilton–Jacobi–Bellman equation for stochastic optimal control problems. In *Proc. 2005 WSEAS International Conference on Dynamical Systems and Control*, pages 489–497, 2005.
- Vladas Pipiras and Murad S Taqqu. *Long-range dependence and self-similarity*, volume 45. Cambridge university press, Cambridge, 2017.
- Jean-Luc Prigent and Philippe Bertrand. Portfolio insurance strategies: a comparison of standard methods when the volatility of the stock is stochastic. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.450061>, 2003.
- Christophe Profeta and Frédéric Vrms. Piecewise constant martingales and lazy clocks. *Probability, Uncertainty and Quantitative Risk*, 4(1):1–27, 2019.
- Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- Amirhossein Sadoghi and Jan Vecer. Optimal liquidation problem in illiquid markets. *European Journal of Operational Research*, 296(3):1050–1066, 2022.
- Enrico Scalas. Five years of continuous-time random walks in econophysics. In *The Complex Networks of Economic Interactions*, pages 3–16. Springer, 2006.
- Enrico Scalas and Bruno Toaldo. Limit theorems for prices of options written on semi-Markov processes. *Theory of Probability and Mathematical Statistics*, 105: 3–33, 2021.
- Enrico Scalas, Rudolf Gorenflo, and Francesco Mainardi. Fractional calculus and continuous-time finance. *Physica A: Statistical Mechanics and its Applications*, 284(1-4):376–384, 2000.

- René L Schilling, Renming Song, and Zoran Vondracek. *Bernstein functions: Theory and Applications*. Berlin, New York: De Gruyter, 2010.
- Jürgen Schmidhuber and Sepp Hochreiter. Long short-term memory. *Neural Computing*, 9(8):1735–1780, 1997.
- L. Schwartz. *Théorie des distributions*. Publ. de l’Inst. de Math. de l’Univ. de Strasbourg. Hermann, 1966.
- Ken Sennewald. Controlled stochastic differential equations under Poisson uncertainty and with unbounded utility. *Journal of Economic Dynamics and Control*, 31(4):1106–1131, 2007.
- Foad Shokrollahi, Adem Kılıçman, and Marcin Magdziarz. Pricing European options and currency options by time changed mixed fractional Brownian motion with transaction costs. *International Journal of Financial Engineering*, 3(01):1650003, 2016.
- Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of computational physics*, 375:1339–1364, 2018.
- Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.
- Aleksander A Stanislavsky. Black–Scholes model under subordination. *Physica A: Statistical Mechanics and its Applications*, 318(3-4):469–474, 2003.
- Harvey J Stein, Albert Cohen, and Nick Costanzino. A unified framework for default modeling. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4098129>, 2022.
- Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- Peter Tankov. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2003.
- Bruno Toaldo. Convolution-type derivatives, hitting-times of subordinators and time-changed C_0 -semigroups. *Potential Analysis*, 42(1):115–140, 2015.
- Nicolo Torre. Barra market impact model handbook. *BARRA Inc., Berkeley*, 208, 1997.
- Lorenzo Torricelli. Trade duration risk in subdiffusive financial models. *Physica A: Statistical Mechanics and its Applications*, 541:123694, 2020.
- Nizar Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media, 2012.

- Peter JM Van Laarhoven and Emile HL Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.
- D Vrabie and Frank L Lewis. Adaptive optimal control algorithm for continuous-time nonlinear systems based on policy iteration. In *2008 47th IEEE Conference on Decision and Control*, pages 73–79. IEEE, 2008.
- Draguna Vrabie, O Pastravanu, Murad Abu-Khalaf, and Frank L Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.
- Jun Wang, Jin-Rong Liang, Long-Jin Lv, Wei-Yuan Qiu, and Fu-Yao Ren. Continuous time Black–Scholes equation with transaction costs in subdiffusive fractional Brownian motion regime. *Physica A: Statistical Mechanics and its Applications*, 391(3):750–759, 2012.
- Paul Werbos. Approximate dynamic programming for real-time control and neural modeling. *Handbook of intelligent control*, 1992.
- Bernard Wong and Charles C Heyde. On changes of measure in stochastic volatility models. *International Journal of Stochastic Analysis*, 2006:1–13, 2006.